

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_p2n model

Run on: October 18, 2004, 09:55:06 ; Search time 18166 Seconds  
(without alignments)  
6226.408 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985  
Sequence: 1 MAGAPPALLPCLISDCC.....RSGVGPSTLPGSHRSPK 3104

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+P2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10016248/runat\_18102004\_105459\_11541/app\_query.faeta\_1.3271  
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=humand40.cdi -LIST=45  
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10016248 @CGN 1\_1\_12209 @runat\_18102004\_105459\_11541 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hc: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_gss1: \*  
9: gb\_gss2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description        |
|------------|--------|-------------|--------|-------|--------------------|
| 1          | 5729   | 33.7        | 6030   | 9     | AY407073 Homo sapi |
| 2          | 5696.5 | 33.5        | 5901   | 9     | AY407075 Mus muscu |
| 3          | 4805.5 | 28.3        | 3882   | 3     | AK081081 Mus muscu |
| 4          | 4001   | 23.6        | 5824   | 9     | AY407074 Pan trogl |
| 5          | 1284   | 7.6         | 748    | 7     | CF744273 UT-M-GV0  |
| 6          | 1229.5 | 7.2         | 2841   | 3     | AK046385 Mus muscu |
| 7          | 1182   | 7.0         | 666    | 7     | CF531622 UI-M-PY0  |
| 8          | 1126.5 | 6.6         | 2809   | 5     | BO832212 C82 APT02 |
| 9          | 1122.5 | 6.6         | 673    | 5     | BX671124 BX671124  |

|    |        |     |      |   |          |                     |
|----|--------|-----|------|---|----------|---------------------|
| 10 | 1113.5 | 6.6 | 733  | 5 | BU370222 | BU370222 603595935  |
| 11 | 1091   | 6.4 | 659  | 2 | BB309178 | BB309178 BB309178   |
| 12 | 1053   | 6.2 | 707  | 7 | CK312578 | CK312578 SB02035A2  |
| 13 | 1011   | 6.0 | 661  | 1 | AL638364 | AL638364 AL638364   |
| 14 | 1000   | 5.9 | 656  | 6 | BY734600 | BY734600 BY734600   |
| 15 | 933    | 5.5 | 671  | 5 | BX676029 | BX676029 BX676029   |
| 16 | 931    | 5.4 | 3496 | 5 | BC032550 | BC032550 Homo sapi  |
| 17 | 911.5  | 5.4 | 3348 | 3 | AK003072 | AK003072 Mus muscu  |
| 18 | 894    | 5.3 | 666  | 4 | BG080336 | BG080336 H3052D08-  |
| 19 | 885    | 5.2 | 686  | 7 | CF742205 | CF742205 UI-M-HB0-  |
| 20 | 884.5  | 5.2 | 803  | 4 | BG217886 | BG217886 RST37610   |
| 21 | 872.5  | 5.1 | 3127 | 6 | CB548140 | CB548140 AMGNNUC: C |
| 22 | 860.5  | 5.1 | 6075 | 3 | AY261899 | AY261899 Ciona int  |
| 23 | 854    | 5.0 | 3809 | 3 | BC032548 | BC032548 Homo sapi  |
| 24 | 850    | 5.0 | 683  | 6 | CD634832 | CD634832 56013038J  |
| 25 | 841.5  | 5.0 | 617  | 2 | BF369877 | BF369877 QV4-GN012  |
| 26 | 837.5  | 4.9 | 792  | 4 | BG215218 | BG215218 RST34882   |
| 27 | 832.5  | 4.9 | 685  | 6 | CD634834 | CD634834 56013054J  |
| 28 | 809    | 4.8 | 624  | 6 | CD634836 | CD634836 56013138J  |
| 29 | 806    | 4.7 | 458  | 1 | AT928999 | AT928999 aus6h10.Y  |
| 30 | 800    | 4.7 | 574  | 6 | CD634867 | CD634867 56035822H  |
| 31 | 799    | 4.7 | 572  | 6 | CD634862 | CD634862 56035730J  |
| 32 | 799    | 4.7 | 574  | 6 | CD634863 | CD634863 56035806H  |
| 33 | 799    | 4.7 | 574  | 6 | CD634869 | CD634869 56035830H  |
| 34 | 799    | 4.7 | 2888 | 9 | AY414540 | AY414540 Homo sapi  |
| 35 | 796    | 4.7 | 4379 | 3 | BC053923 | BC053923 Mus muscu  |
| 36 | 795    | 4.7 | 574  | 6 | CD634865 | CD634865 56035814H  |
| 37 | 790    | 4.7 | 569  | 6 | CD634866 | CD634866 56035814J  |
| 38 | 789.5  | 4.6 | 668  | 6 | CD634821 | CD634821 56002879H  |
| 39 | 789.5  | 4.6 | 717  | 6 | CD634825 | CD634825 56002895H  |
| 40 | 788    | 4.6 | 574  | 6 | CD634859 | CD634859 56035722H  |
| 41 | 780    | 4.6 | 693  | 6 | CD634827 | CD634827 56002995H  |
| 42 | 776.5  | 4.6 | 708  | 6 | CD634837 | CD634837 56013154H  |
| 43 | 776.5  | 4.6 | 3663 | 3 | AK083229 | AK083229 Mus muscu  |
| 44 | 773    | 4.6 | 568  | 6 | CD634870 | CD634870 56035830J  |
| 45 | 772    | 4.5 | 3388 | 3 | AK004995 | AK004995 Mus muscu  |

#### ALIGNMENTS

|            |   |             |     |        |                 |
|------------|---|-------------|-----|--------|-----------------|
| RESULT 1   | AY407073  | 6030 bp     | DNA | linear | GSS 15-DEC-2003 |
| LOCUS      | AY407073  |             |     |        |                 |
| DEFINITION | Homo sapiens CSMD1 gene, VIRTUAL TRANSCRIPT, partial sequence.  |             |     |        |                 |
| ACCESSION  | AY407073  |             |     |        |                 |
| VERSION    | AY407073.1  | GI:39763044 |     |        |                 |
| KEYWORDS   | GSS.  |             |     |        |                 |
| SOURCE     | Homo sapiens (human)  |             |     |        |                 |
| ORGANISM   | Homo sapiens  |             |     |        |                 |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.   |             |     |        |                 |
| AUTHORS    | Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriterra, S., Wang, G., Zheng, X.H., White, T.J., Sniinsky, J.J., Adams, M.D. and Cargill, M. |             |     |        |                 |
| TITLE      | Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriterra, S., Wang, G., Zheng, X.H., White, T.J., Sniinsky, J.J., Adams, M.D. and Cargill, M.   |             |     |        |                 |
| JOURNAL    | Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA   |             |     |        |                 |
| COMMENT    | This sequence was made by sequencing genomic exons and ordering them based on alignment.  |             |     |        |                 |
| FEATURES   | Location/Qualifiers   |             |     |        |                 |
| SOURCE     | 1..6030   |             |     |        |                 |

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..26030  
/gene="CSMD1"  
/locus\_tag="HCM2769"

## ORIGIN

## Alignment Scores:

| Pred. No.:             | Score: | Length: | Matches: | Conservative: | Mismatches: | Indels: | Gaps: |
|------------------------|--------|---------|----------|---------------|-------------|---------|-------|
| 5729.00                | 6030   | 6030    | 1008     | 259           | 505         | 59      | 1     |
| Percent Similarity:    | 69.20% |         |          |               |             |         |       |
| Best Local Similarity: | 55.05% |         |          |               |             |         |       |
| Query Match:           | 33.73% |         |          |               |             |         |       |

US-10-016-248-2 (1-3104) x AY407073 (1-6030)

```

QY      1251 LeuGluProGlyTYrAaPheLeuHisIeTYrAaPGLYArgAspSerLeuSerProLeu 1270
DB      1 ATGGAGCCCACTATGACTTCTTACACATCTATGAGGGGAAAGATTCCAAACGCCCTTC 60
QY      1271 ILeGlySerPheTYrGlySerGlnLeuProGlyATgllleGluSerSerAsnSerleu 1290
DB      61 ATTGGAGTTTACAGAGGCTCTCAGGCCCCAGAAAGAAATGAGAGTACGGAACAGCCTG 120
QY      1291 PheLeuAlaPheArgSerAspAlaSerValSerAenAlaGlyPheValIleAspTYrThr 1310
DB      121 TTTCGTGGCATTTGGAGTGAATGCTCCGTGGGCTTTCAAGGTTCCCATTTAAATTAAA 180
QY      1311 GluAenProArgGluSerCysPheAspProGlySerIleLysAanglyTYrArgValGly 1330
DB      181 GNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCAGAAATATATGATGGGACAAAGATTGGA 240
QY      1331 SerAspLeuLysLeuGlySerSerValThrTYrCysHisGlyTYrGluValGlu 1350
DB      241 ACAGACTTCAGCTTGCTGCTCAGCATCACTACCAAGTGTGACTTGSCCTATTAAGATTCTT 300
QY      1351 GlyThrSerThrLeuSerCysIleLeuGlyProAspGlyLysProValITrAaenPro 1370
DB      301 GACCCCTATCATCATCAGCTGTGTGATGGGGCTGATGGAAACCTCTGGACCAAGT 360
QY      1371 ArgProValCysThrAlaProCysGlyGlyGlnTYrValIleLysSerAspGlyValIleu 1390
DB      361 CTGCGCTTCCTGATGCTCTCTGTGAGGCGAGACCGGATCAAGAGGGGTAGTTTAA 420
QY      1391 SerProAenTYrProGlnAenTYrThrSerGlyGlnIleCysLeuTYrPheValThrVal 1410
DB      421 TCACCAAACTAACCCCATTAATTACACAGCTGTCAAAATATGCTCTATTCATCAGCGTA 480
QY      1411 ProLysAspTYrValValPheGlyGlnPheAlaPhePheHisThrAlaLeuAenAspVal 1430
DB      481 CCAAGAGAAATTCGNNNNNNNNNNNNNNNNNNNNNNNNNNNGACACCTGATGATTGG 540
QY      1431 ValGluValHisArgGlyHisSerGlnHisSerArgLeuLeuSerSerLeuSerGlySer 1450
DB      541 GCAGAAATTTATGATGAGAACCATGCAAGCGCACTTCTGACTCAGCTCTGGGGTCT 600
QY      1451 HisThrGlyGluSerleuProLeuAlaThrSerAenGlnValleuIleLysPheSerAla 1470
DB      601 CACTCAGGGGAAACATTGCTTGTGCTAGCTCAAAATCAATCTGCTCGATTGAGTCA 660
QY      1471 LysGlyLeuAlaProAlaArgGlyPheHisPheValTYrGlnAlaValProArgThrSer 1490
DB      661 AAGAGGGGTGCTCTCCCGCGGCTTCCACTTGATCAAGCTGTTCTCGTCAACCGT 720
QY      1491 AlaThrGlnCysSerSerValProGluProArgTYrGlyLysArgLeuGlySerAspPhe 1510
DB      721 GACACCAATGACAGCTCTGCTCCCGAGCCCAATGACGAGAGAGAAATTGTTCTGAGTTT 780
QY      1511 SerValGlyAlaIleValArgPheGlyCysAenSerGlyTYrAlaLeuGlnGlySerPro 1530
DB      781 TTTCGGCGGTCCATGCTCGATTGAGTGCAGACCCGGGATACCTGCTTCAAGGTTCCACG 840

```

```

QY      1531 GluIleGlyCysLeuProValProGlyAlaLeuAlaGlnTrpAenValSerAlaProThr 1550
DB      841 GCGCTCCACTGCGCAGTCCGTCCGCAACGCTTGGCAACAGTGGAAACAGCATGCCACAG 900
QY      1551 CysValValProCysGlyGlyAenLeuThnGluArgArgGlyTYrTrileuSerProGly 1570
DB      901 TGTGTGGTACCTGCAAGTGGCAATTCATCTCAACGAAAGGTACAAATCTGTCCCGGCG 960
QY      1571 PheProGluProTYrLeuAenSerLeuAenCysValITrPylsIleValValProGluGly 1590
DB      961 TACCTGAGGCAATACGAAACAACTTGAATCTGATGAGAGATCATAGTTACGAGGCG 1020
QY      1591 AlaGlyIleGlnIleGlnValValSerPheValThnGlnAAsnTrpAspSerLeuGlu 1610
DB      1021 TCGGAATTCAGATTCAGATGATGATTTGGCAGAGAGCAAGACTGGACTCTCTTGAG 1080
QY      1611 ValPheAspGlyAlaAspAenThrValThrMetLeuGlySerPheSerGlyTYrThrVal 1630
DB      1081 ATCCAGATGGTGGGATGTGACCCGACCACTGGGAGCTTCTGAGNNNNNNNNNN 1140
QY      1631 ProAlaLeuAenSerThrSerAenGlnLeuTYrLeuHisPheTYrSerAspIleSer 1650
DB      1141 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1200
QY      1651 ValSerAlaIleGlyPheHisLeuGluTYrIleThrValGlyLeuSerCysProGlu 1670
DB      1201 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1260
QY      1671 ProAlaValProSerAspGlyValIleThrGlyGluArgTYrLeuValAsnAspValVal 1690
DB      1261 CCAGCCTCTCCAGCAAGATCAAAATCGAATCGTCAATGAGTGAAGAGAGTGTCTC 1320
QY      1691 SerPheGlnCysGluProGlyTYrAlaLeuGlnGlyHisAlaHisIleSerCysMetPro 1710
DB      1321 TCCTTCCAGTCCAGGCCGCGGTACACCTGAGAGGCCGCTTCCCAATTTCTGTATGCCA 1380
QY      1711 GlyThrValAlaGlyTrpAenTYrProProLeuCysIleAlaGlnCysGlyGlyTYr 1730
DB      1381 GGGACCTGTGCGGTGGAACTATCCGTCCTCTGTGCAATTCGATGGAGGGACG 1440
QY      1731 ValGluGluMetGluGlyValIleLeuSerProGlyPheProGlyAenTYrProSerAen 1750
DB      1441 CTGAGACCTTGGGTGTGATGATCTCAGCCCGGCTTCCAGGTTTCAACCCCAACAC 1500
QY      1751 MetAspCysSerTrpLysIleAlaLeuProValGlyPheGlyAlaHisIleGlnPheLeu 1770
DB      1501 TTAGACTGCACTGGAGATCTCATTCACATCGGCTATGTCACATATTCAGTTCTG 1560
QY      1771 AsnPheSerThrGluProAenHisAspTYrIleGluIleArgAenGlyProTYrGluThr 1790
DB      1561 AATTTTCTACCAAGGATTAATCAATGACTTCTTGAATTAATCAATGACCTTACCAAC 1620
QY      1791 SerArgMetGlyArgPheSerGlySerGluLeuProSerSerLeuSerLeuSerTrpSer 1810
DB      1621 AGCCCACTGATGAGCAATTTAGCGGACGAGATCTCCCGCGCTCTGTGAGCAACAG 1680
QY      1811 HisGluThrThrValTYrPheHisSerAspHisSerGlnAsnArgProGlyPheLysLeu 1830
DB      1681 CATGAAACCTTCACCACTTTATATGACCAATTCGCAAAACCGGACAGATTTAACTT 1740
QY      1831 GluTYrGlnAlaTYrGluLeuGlnGlyCysProAspProGluProPheAlaAenGlyIle 1850
DB      1741 GCTTACCAAGCCATTAATTAAGAACTGTCCAGATCCACCCCATTTAGATGAGTATC 1800
QY      1851 ValArgGlyAlaGlyTYrAenValGlyGlnSerValThrPheGluCysLeuProGlyTYr 1870
DB      1801 ATGATCAACTCGGATTAACGCTGGGGCAATCAGATATCTTGAGATGTTATCTGGGTAC 1860
QY      1871 GlnLeuThrGlyHisProValLeuThrCysGlnHisGlyTYrAenArgAenTrpAspHis 1890
DB      1861 ATTCTAATAGGCAATCTGTCTCTGATGACATGGAGATCAAGAAATCGGAATTAAC 1920

```

|    |      |  |      |
|----|------|--|------|
| OY | 1891 | ProleuProlybCybGluValProCybGlyValAsnIleThrSerSerAsnGlyThrVal   | 1910 |
| Db | 1921 | CTTTTCCAAAGATGATGCCCTTGTGGGATCAACGTAACCTTCCAGAACGCCATC         | 1980 |
| OY | 1911 | TySerProGlyPheProSerProTyzSerSerSerGlnAspCyValTrpLeuIleThr     | 1930 |
| Db | 1981 | TACTCCCTGGCTTCTCTGATGAGTATCCGATCCGAAAGACTGATTTGGCTATCAG        | 2040 |
| OY | 1931 | ValProIleGlyHisGlyValArgLeuAsnLeuSerIleLeuGlnThrGluProSerGly   | 1950 |
| Db | 2041 | GTGCCTCAGGGCAGAGATTATCAATCAACTTCAACCTGTTACAGACGAAGCTGTCAAC     | 2100 |
| OY | 1951 | AspHeIleThrIleTrpAspGlyProGlnGlnThrAlaProArgLeuGlyValPheThr    | 1970 |
| Db | 2101 | GATTACATTCGTGTGGAGCGGTCCCATCAAGAACCCAGCTGGAGTTTCAGT            | 2160 |
| OY | 1971 | ArgSerMetAlaIleValbThrValGlnSerSerSerAsnGlnValLeuLeuPheHis     | 1990 |
| Db | 2151 | GGCAACACAGCCCTCGAAGCGCGATATAGCTTCCACACCAAGTCTCTCAAGTTCCAC      | 2220 |
| OY | 1991 | ArgAspAlaAlaThrGlyIlePheAlaIleAlaPheSerAlaTyProLeuThrIlys      | 2010 |
| Db | 2221 | AGCGACTTTCCAAATGGAGGGCTTCTTGTCCTCAATTTCCAGCATTTCACTCAAGAA      | 2280 |
| OY | 2011 | CysProPProProThrIleLeuProAsnAlaGluValValThrGluAsnGluPheAsn     | 2030 |
| Db | 2281 | TGTCAACCTCCCCCGAGCTTCCACGACGAAAGCTTACTCGAGATGATGATTTCCAG       | 2340 |
| OY | 2031 | IleGlyAspIleValArgTyArgCysLeuProGlyPheThrLeuValGlyAsnGluIle    | 2050 |
| Db | 2341 | ATAGGAGATTTGTGATGATGACAGGCCACCCCGGGATACCTTGTTGGGAGCGACATT      | 2400 |
| OY | 2051 | LeuThrCysIleValGlyThrTyIleGlnPheGluGlyProProProIleCysGluVal    | 2070 |
| Db | 2401 | CTGACTTCGAAGCTCAGTTCACAGTTGAGTTAGCGTTCCTCCCAACATGTGAAACA       | 2460 |
| OY | 2071 | HisCysProThrAsnGluLeuLeuThrAspSerThrGlyValIleLeuSerGlnSerTy    | 2090 |
| Db | 2461 | CAATGCCACGAAGAAAGAGCTCGGACTGATCATCGGAGATCATTTCTCAGTCCAGGAT     | 2520 |
| OY | 2091 | ProGlySerTyPProGlnPheGlnThrCysSerThrPLeuValArgValGluProAspTy   | 2110 |
| Db | 2521 | CGGGGATATATTTTAACTCCAGACTGTGCTTGAGATTAATAAGTGAACAAACATAC       | 2580 |
| OY | 2111 | AsnIleSerLeuThrValGluTyArgPheLeuSerGluTyArgGlyAspGluPheGluIle  | 2130 |
| Db | 2581 | AACATTACATCTTGTGACACATTTCAAACTGAAAGCAGTTTGATGACTGGAACTG        | 2640 |
| OY | 2131 | PheAspGlyProSerGlyGlnSerProLeuLeuValAlaLeuSerGlyAsnTySerAla    | 2150 |
| Db | 2641 | TTTGTATGTTCTTCTGGGCAAAAGTCTCTGCTAGTAAAGCTTAAGGGGAATCATATCAA    | 2700 |
| OY | 2151 | ProleuIleValThrSerSerSerAsnSerValTyIleAsnGlyTrpSerSerAspHisAla | 2170 |
| Db | 2701 | CAATCAAAATTTTACAGACGAGAGATATAGTTAATCTCCGCTGGTCCACATGACATAGCC   | 2760 |
| OY | 2171 | TyrAsnArgIleGlyPheIleGlyIleArgTySerAlaProTyArgSerLeuProArgAla  | 2190 |
| Db | 2761 | ACCAAGTAAGAAAGATTCAAAGTTCGCTATGACAGACTTACTGACGTTTAAACCACCC     | 2820 |
| OY | 2191 | ProleuHisGlyPheIleLeuGlyGlnThrSerThrGlnProGlyGlySerIleHisPhe   | 2210 |
| Db | 2821 | CTGAAGAATGGGGGATATCTTAAACAAGACTCAGAGAGCGTTGGAAACGAAGTCATATT    | 2880 |
| OY | 2211 | GlyCysAsnAlaGlyTyArgLeuValGlyHisSerMetAlaIleCysThrArgHisPro    | 2230 |
| Db | 2881 | TTTTCGAAGCCTGATATACCAGATGCTGGCCACAGCAATGTGAACCTGTGAAGCAACCA    | 2940 |
| OY | 2231 | GlnGlyTyThrIleLeuTrpSerGlyAlaIleProLeuCysGlnAlaLeuSerCysGlyLeu | 2250 |
| Db | 2941 | CTTGGCATGTACACAGTGGACTCCCTCAGCCACTCTGCCAGCGCTGTCTCTGTGGATC     | 3000 |
| OY | 2251 | ProGluAlaProIleAsnGlyMetValbPheGlyIleGlyTyThrValGlyThrIleVala  | 2270 |

|    |      |   |      |
|----|------|---|------|
| Db | 3001 | CCAGATCCCGAGAAACGGTTCAATTACCGGGAGACGATTCATTGGACAGTAAAGTG        | 3060 |
| Qy | 2271 | ValTyrSerCysSerGluGlyTyrHisLeuGlnIaGlyIaGlyIaThrAlaGlyCys       | 2290 |
| Db | 3061 | GTCTATGAATGTACTAGAGCGCTTCAAGCTTGAAATCCAGCCAGACAGAACAGCCGTGT     | 3120 |
| Qy | 2291 | LeuAspThrGlyLeuTyrTrpSerAsnArgAsnValProProGlnCysValProValThrCys | 2310 |
| Db | 3121 | CAAGAAAGATGGCGCTGTGGAGTAAACAAAGGGAGCCGCCATGTGTAAAGCCGGTGCCTTGC  | 3180 |
| Qy | 2311 | ProAspValSerSerIleSerValGlnHisGlyArgTyrPargLeuIlePheGluThrGln   | 2330 |
| Db | 3181 | CCAGCATTTGAAGTCAAGCTCTCAGAACATGTCAATCTGGAGGCGGTTCAGAGATCTTG     | 3240 |
| Qy | 2331 | TyrGlnPheGlnIaGlnLeuMetLeuIleCysAspProGlyTyrTyrTyrThrGlyGln     | 2350 |
| Db | 3241 | AATAGATACGGTGTCTCAAGATTTCTCTAGCTGACGTCTCTGTACTTAAAGAGCTGG       | 3300 |
| Qy | 2351 | ArgValIleArgCysGlnIaAsnGlyLysTrpSerLeuGlyAspSerThrProThrCys     | 2370 |
| Db | 3301 | AGGCTCTCGCGTCCAGGCCAAATGGAGCGGTGAACATAGAGATGTAGAGGCCAAGCTGT     | 3360 |
| Qy | 2371 | ArgIleIleSerCysGlyGluLeuProIleProAsnGlnHisArgIleGlyThrLeu       | 2390 |
| Db | 3361 | CGA-----  | 3363 |
| Qy | 2391 | SerValTyrGlyAlaThrAlaIlePheSerCysAsnSerGlyTyrThrLeuValGlySer    | 2410 |
| Db | 3363 | -----   | 3363 |
| Qy | 2411 | ArgValArgGlyCysMetAlaAsnGlyLeuTrpSerGlySerGluValArgCysLeuAla    | 2430 |
| Db | 3364 | -----GCT  | 3366 |
| Qy | 2431 | GlyHisCysGlyThrProGluProIleValAsnGlyHisIleAsnGlyGluAsnTyrSer    | 2450 |
| Db | 3367 | GGCCACGCGGTTCCCGAACCCCATTTGTAAACGCTCAATTGTGGAGATGGCTTCAGT       | 3426 |
| Qy | 2451 | TyrArgGlySerValAlaTyrGlnCysAsnAlaGlyPheArgLeuIleGlyMetSerVal    | 2470 |
| Db | 3427 | TACAGACACAGCGGTGTTTACCACTGCAATCTCGTTCGCGCTGTGGAGAACTCCGTG       | 3486 |
| Qy | 2471 | ArgIleCysGlnGlnAspHisHisTrpSerGlyLysThrProPheCysValProIleThr    | 2490 |
| Db | 3487 | AGGATATGCCCGCAAGACCAACAAGTGTCTGGACAAACGCGTGTCTGTGCCATCACA       | 3546 |
| Qy | 2491 | CysGlyHisProGlyAsnProValAsnGlyLeuThrGlnGlyAsnGlnPheAsnLeuAsn    | 2510 |
| Db | 3547 | TGTGTATCACCTCGAAGAACCTCGCCACAGGATTCCTATATGGCAGTGAAGTTCAACCTGAAT | 3606 |
| Qy | 2511 | AspValValLysPheValCysAsnProGlyTyrPheAlaGlyGluAlaIaIaArgSerGln   | 2530 |
| Db | 3607 | GATGTCGGAAATTTCACCTGCAACACGGGCTATTGTGTGCGAGGCGTGTCCAGGCCAG      | 3666 |
| Qy | 2531 | CysLeuAlaSerGlyGlnTrpSerAspMetLeuProThrCysArgIleIleAsnCysThr    | 2550 |
| Db | 3667 | TGTGGAGAACACGGCCAGTGGAGTAAAGCCCTGTGCCACGtGTGGAGTGGAACTGTCT      | 3726 |
| Qy | 2551 | AspProGlyHisGlnGluAsnSerValArgGlnValHisAlaSerGlyProHisAspPhe    | 2570 |
| Db | 3727 | GATCCAGGCTTTTGGAGAAATGCAATTCGTGCACGGCGCAACAGAACTTCCCTGAGAGTTT   | 3786 |
| Qy | 2571 | SerPheGlyThrThrValSerTyrArgCysAsnHisGlyPheTyrLeuLeuGlyThrPro    | 2590 |
| Db | 3787 | GAGTATGATAGATATCTGTACATTCGACAGAGAGGAAATTTTAACTTGTGGAAATCTTCA    | 3846 |
| Qy | 2591 | ValLeuSerCysGlnGlyAspGlyThrTrpAspArgProArgProGlnCysLeuLeuVal    | 2610 |
| Db | 3847 | GCGTTGAACCTGTATGGCAATAGCTTATGGGACCGAATCCCTGCCAAGTCTTTTGGCATATA  | 3906 |
| Qy | 2611 | SerCysGlyHisProGlySerProProHisSerGlnMetSerGlyLysSerTyrThrVal    | 2630 |

Db 3907 TCCTGTGACACCCAGGAGTCCCTGCAACGCCGTCTCCTGAGAGCTGTTTACTAT 3966  
 QY 2631 GYVALValValArgTyrSerCysIleGlyValArgThrLeuValGlyAsnSerThrArg 2650  
 Db 3967 GGGCGCGTGTGACACTTCTCTGACAGAGGAGCGAGGAGCTCATAGGACACGACGAGA 4026  
 QY 2651 MetCysGlyLeuAspGlyHisTyrThrGlySerLeuProHisCysSerGlyThrSerVal 2670  
 Db 4027 GTGTGACAGAGAGACAGTCACTGAGCGGGGCACTGCCCACTGACACGAAATATCTCT 4086  
 QY 2671 GlyValCysGlyAspProGlyIleProAlaHisGlyIleArgLeuGlyAspSerPheAsp 2690  
 Db 4087 GGAATCTGTGTGATCCGGGAGCCGACGACATGGGCTCGGGCTTGGTGAATGACTTAA 4146  
 QY 2691 ProGlyThrValMetArgPheSerCysGlyAlaGlyHisValLeuArgGlySerSerGly 2710  
 Db 4147 ACAAGAGATCTTCTCGCTTCTCTGTGAATGGGGCCACGAGGGGGCTCCCTGAA 4206  
 QY 2711 ArgThrCysGlyAlaAsnGlySerTyrPserGlySerGlnProGlyCysGlyValIleSer 2730  
 Db 4207 CGCAGGTGTGGCTCAATGGGTGATGTCAGGACTGACGCGGTGTGAGCGGTCTCC 4266  
 QY 2731 CysGlyAspProGlyThrProSerAsnAlaArgValValPheSerAspGlyLeuValPhe 2750  
 Db 4267 TGTGCAACCTCTGGACACCCACACGAAATGATGTGACTGATGATGGCAATCTGTTTC 4326  
 QY 2751 SerSerSerIleValTyrGlyCysArgGlyIleTyrTyrAlaThrGlyLeuLeuSerArg 2770  
 Db 4327 TCCAGCTGTGATCATCTATGCTGTGGAGAGCTACAAAGCTCAGGGCTCATGACACGG 4386  
 QY 2771 HisCysSerValAsnGlyThrTyrThrGlySerAspProGlyCysLeuValIleAsnGly 2790  
 Db 4387 CATTCGACAGCCCAATGGAGCTGAGCAGGCACTGCTCCCACTGCAACAATTAATGTTTC 4446  
 QY 2791 GlyAspProGlyIleProAlaAsnGlyLeuArgLeuGlyAsnAspPheArgTyrAsnIle 2810  
 Db 4447 GGGGATCAGGCAACATAGGAAATGGATCATCACTTTGGACCACTTCACTTCAACAG 4506  
 QY 2811 ThrValThrTyrGlnCysValProGlyTyrMetMetGlnSerHisArgValSerValLeu 2830  
 Db 4507 ACTGTGAGCTATCATGTAACCCAGGCTATGTATGTAAGACATGACATCCGCTACTT 4566  
 QY 2831 SerCysThrIleAspArgTyrThrTPAsnGlyThrIleProValCysIleValMetCys 2850  
 Db 4567 CGCGTACCAAGACCGGAGAGTGAATCCAGCAACCTGTCTGCAAGCGGCTGTCT 4626  
 QY 2851 LysProProProLeuIleProAsnGlyLysValValGlySerAspPheMetTyrGlySer 2870  
 Db 4627 CCTCAGCGCGCGCGGAGAGATGGAACAGTGAAGAGTATTTCCGCTGGGCTCC 4686  
 QY 2871 SerValThrTyrAlaCysIleGlnGlyIleLeuProGlnCysPheProValPheThrCys 2890  
 Db 4687 AGCATTAAGTAAAGCTGATGAGCGGTACCACTTCTCCTCCCTCCCTCTCTCTCT 4746  
 QY 2891 GlnGlyAsnGlySerTyrThrGlyIleLeuProGlnCysPheProValPheCysGlyAsp 2910  
 Db 4747 GAAGGTGCGGGGTGTGGAAGAGAGATCCCACTGTCTCTCTGTCTCTGGAGAGC 4806  
 QY 2911 ProGlyValProSerArgGlyArgGlyAspArgGlyPheSerTyrArgSerSerVal 2930  
 Db 4807 CCTGGCATCCCGCAGAAAGGCGACTTATGGGAAAGTTTCACTATTAAGCCGAAAGTC 4866  
 QY 2931 SerPheSerCysHisProProLeuValLeuValGlySerProArgArgPheCysGlnSer 2950  
 Db 4867 TTCTTCAGTGAACATCTCATTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4926  
 QY 2951 AspGlyThrTyrPserGlyThrGlnProSerCysIleAspProThrLeuThrCysAla 2970  
 Db 4927 NNN 4986  
 QY 2971 AspProGlyValProGlnPheGlyIleGlnAsnAsnSerGlnGlyTyrGlnValGlySer 2990  
 Db 4987 NNN 5046

QY 2991 ThrValLeuPheArgCysGlnIleGlyTyrLeuLeuGlnGlySerThrThrArgThrCys 3010  
 Db 5047 NNN 5106  
 QY 3011 LeuProAsnLeuThrTyrPserGlyThrProProAspCysValProHisCysArgGln 3030  
 Db 5107 NNN 5166  
 QY 3031 ProGlyThrProThrHisAlaAsnValGlyAlaLeuAspLeuProSerMetGlyTyrThr 3050  
 Db 5167 CAGAAACCCCGGACACCGGATGTAGAGCCATGATCTTCTTCACTTTCGCTACACC 5226  
 QY 3051 LeuIle-ThrProAlaArgArgAlaSerProSerArgValAlaProSerThrAlaProAl 3070  
 Db 5227 TTAGTGTACCTGTCATCCATCCTTTTCTCTGCGAGGGGATCTGAGACAGAACATCT 5286  
 QY 3070 AARGArgMetAlaAlaGlyGlnAlaSerArg 3080  
 Db 5287 AAGACAGACATGAATGACAGAAAGTCGC 5317  
 RESULT 2  
 AY407075 5901 bp DNA linear GSS 15-DEC-2003  
 LOCUS AY407075 Mus musculus CSM01 gene, VIRUAL TRANSCRIPT, partial sequence,  
 DEFINITION genomic survey sequence.  
 ACCESSION AY407075 GI:39763046  
 VERSION AY407075.1 GI:39763046  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 5901)  
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,  
 Todd, M.A., Tenenbaum, D.M., Clevello, D.R., Lu, F., Murphy, B.,  
 Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Smitsky, J.J.,  
 Adams, M.D. and Cargill, M.  
 Direct Submission  
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA  
 This sequence was made by sequencing genomic exons and ordering  
 them based on alignment  
 COMMENT location/Qualifiers  
 FEATURES  
 source  
 1..5901  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 <1..>5901  
 /gene="CSMD1"  
 /locus\_tag="HMC2769"  
 gene  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 5901  
 Score: 5696.50 Matches: 1014  
 Percent Similarity: 69.02% Conservative: 249  
 Best Local Similarity: 55.41% Mismatches: 467  
 Query Match: 33.54% Indels: 103  
 DB: 9 Gaps: 4  
 US-10-016-248-2 (1-3104) x AY407075 (1-5901)  
 QY 1251 LeuGluProGlyTyrAspPheLeuHisIleTyrAspGlyArgAspSerLeuSerProLeu 1270

```

Db      1 ATGAGCGCGAGTTGACTTCTCCATATATAGAGGAGGAGACTCCACAGCCCACTG 60
Qy      116GlySerPheTyrglySerGlnLeuProGlyArgIleGlySerSerSerSerLeu 1230
Db      61 ATGGAGGTTTCCAGGGCTCTCAAGCCCAAGAGAGATCGAGAGCATGTGTAACAGCTC 120
Qy      1291 PheLeuAlaPheArgSerAspAlaSerValSerAsnAlaGlyPheValIleAspTyrThr 1310
Db      121 TTTCGTGGCATTCAGAGATGATGCTTCTGTGGCCGTGGGATTCGCCATTGAATTAA 180
Qy      1311 GlnAsnProAlaGlySerCysPheAspProGlySerIleValAsnGlyTyrArgValGly 1330
Db      181 GAGAAACCCCGGAGAGCTTCTTGACCCCTGGGAAACATAAGATGGAGACAAGATTGGA 240
Qy      1331 SerAspLeuIleuGlySerSerValThrTyrCysHisGlyValTyrGlnValGlu 1350
Db      241 ACTGACTTTAACTGAGGCTCTACAGTTACCTATCAGTGTGACTCTGTGTACAAAGATTGTG 300
Qy      1351 GlyThrSerThrLeuSerCysAlaLeuGlyProAspGlyLysProValTyrAsnAspPro 1370
Db      301 GATCCCTGCTCATTAAGTGTGTGACAGAGGCTGATGGAGCCGCTCTGSCA----- 353
Qy      1371 ArgProValCysThrAlaProCysGlyGlyGlnTyrValGlySerAspGlyValIleu 1390
Db      354 -----CCCTGTGAGGCGCAGTACACGGGCTCGAGGGGAGTATTG 395
Qy      1391 SerProAspTyrProGlnAsnTyrThrSerGlyGlnIleCysLeuTyrPheValThrVal 1410
Db      396 TCACCAACATACCTCCATTAACCTACACAGTGGAGAGATGTCACTTATTCATACAGCTG 455
Qy      1411 ProLysAspTyrValIleValPheGlyGlnPheAlaPhePheHisThrAlaLeuAsnAspVal 1430
Db      456 CCCAAGGAATTCGTGTGTTTGGACAGTTTCCATTTCACAGATGCACATAAGACTTG 515
Qy      1431 ValGluValHisAspGlyHisSerGlnHisSerArgLeuLeuSerSerLeuSerGlySer 1450
Db      516 GCAGAGTTGTTGTGGAACACATCCCGACGCGAGCTTGT----- 556
Qy      1451 HisThrGlyGlnSerLeuProLeuAlaThrSerArgGlnValIleLeuLysPheSerAla 1470
Db      557 -----GAAACACTGCACATGAGTACATCAATCAGATTCGCTTCGATTACAGGCC 607
Qy      1471 LysGlyLeuAlaProAlaArgGlyPheHisPheValTyrGlnAlaValProArgThrSer 1490
Db      608 AAAAGCGAGCTTCTGCACGSGGTTTCCACTTCTCTCAAGCTGTTCCGCGACCACT 667
Qy      1491 AlaThrGlnCysSerSerValProGluProArgTyrGlyLysArgLeuGlySerAspPhe 1510
Db      668 GATACACAGTGCAGCTGTCCCGAGGCCAGATATGGAGAGAAGATTGTTGAGATTTC 727
Qy      1511 SerValGlyAlaIleValArgPheGlnCysAsnSerGlyTyrAlaLeuGlnGlySerPro 1530
Db      728 TCCCAGAGCTTCATCGTCCGCTCGAGTGTAAACCAGGTTACTCTGTCGAAAGTTCACAC 787
Qy      1531 GluIleGlyCysLeuProValProGlyValaLeuAlaGlnTyrAsnValSerAlaProThr 1550
Db      788 GCCATCCGCTGCAGCTGTGCCAAATGCTTGGCTCAGTAAGAACACACATCCCAAGC 847
Qy      1551 CysValValProCysGlyGlyAsnLeuThrGlnArgArgGlyTyrIleLeuSerProGly 1570
Db      848 TGTTAGTTCATCAGAGGCACTTCACTCAGAGAAAGAGAGGATTTTCTCCCAAGC 907
Qy      1571 PheProGluProTyrLeuAsnSerLeuAsnCysValTyrIleValValProGluGly 1590
Db      908 TATCTTGAGCCCTATGGAGAACCTGTAACTGTGTGTGGAAGATCATGTATCCGAGGCG 967
Qy      1591 AlaGlyIleGlnIleGlnValIleSerPheValThrGlnIleAsnTyrAspSerLeuGlu 1610
Db      968 TCGGGATTCAGATTCAGATGATGACTTTCACAGAGCAAGAACTGGGACTTCCCTGAG 1027
Qy      1611 ValPheAspGlyAlaAspAsnThrValThrMetLeuGlySerPheSerGlyThrThrVal 1630

```

---

```

Db      1028 ATCCATGATGAGAGACATGACGCGCCCGAGACTGCGGACGCTTCTCAANNNNNNNN 1087
Qy      1631 ProAlaLeuLeuAsnSerThrSerAsnGlnLeuTyrLeuHisPheTyrSerAspIleSer 1650
Db      1088 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1147
Qy      1651 ValSerAlaAlaGlyPheHisLeuGluTyrLysThrValGlyLeuSerSerCysProGlu 1670
Db      1148 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1207
Qy      1671 ProAlaValProSerAsnGlyValLysThrGlyGlnArgTyrLeuValaAsnValVal 1690
Db      1208 CCTGCTCTCCCGACCAATGCGATCAAGATGAGAGACCGGATCAATGTAAGATGTCTG 1267
Qy      1691 SerPheGlnCysGluProGlyTyrAlaLeuGlnGlyHisAlaHisIleSerCysMetPro 1710
Db      1268 TCTTTCAGTGTAACTGGGATTACTTGTGAGGCGCGCTCACACATTTCTTGATGCCA 1327
Qy      1711 GlyThrValArgArgTyrAsnTyrProProLeuGlyAlaGlnCysGlyGlyThr 1730
Db      1328 GGAACCGTACGTGCTGGAACCTACCTCCCTTGTGATGTCACCTGTGGTGGGACA 1387
Qy      1731 ValGluIleuSerGlyValIleLeuSerProGlyPheProGlyAsnTyrProSerAsn 1750
Db      1388 CTAACCAAGATGATGAGATGATCTTGAGCCAGAGGTTCCAGGCTCATACCCCAACAC 1447
Qy      1751 MetAspCysSerTyrLysIleAlaLeuProValGlyPheGlyAlaHisIleGlnPheLeu 1770
Db      1448 TTGAGCTGACCTGGAGATATCACTGCCCTTGTGATGTGACATATCCAAATTTCTA 1507
Qy      1771 AsnPheSerThrGluProAsnHisAspTyrIleGluIleArgAsnGlyProTyrGluThr 1790
Db      1508 AATTTCATCTGAAGCAACACATCACTGAGATCCAGAAATGAGCCCTTACACAGC 1567
Qy      1791 SerArgMetCysGlyArgPheSerGlySerGluLeuProSerSerLeuLeuSerThrSer 1810
Db      1568 AGTCCAAATGATGGGCGAGTTCAGTGGCCCTGACCTGCTCATATGTCGAGACACACA 1627
Qy      1811 HisGluThrThrValTyrPheHisSerAspHisSerGlnAsnAspProGlyPheLysLeu 1830
Db      1628 CACGAAACCTCATCCGCTCTATAGTACCACTCACAGAAACCGGCAAGATTAACTC 1687
Qy      1831 GluTyrGlnAlaTyrGlnLeuGlnGlnCysProAspProGluProPheAlaAsnGlyIle 1850
Db      1688 AGTTACCAAGCTTATGATGATACAGAACTGCCAGATCCACCCGATTCAGAAATGGGTTTC 1747
Qy      1851 ValArgIleAlaGlyTyrAsnValGlyGlnSerValThrPheGluCysLeuProGlyTyr 1870
Db      1748 ATGATCAACTCGGATTCACGCTGGGCGCAGTCCGATCTCTTGTGAGTGTATCCCGGCTAC 1807
Qy      1871 GlnLeuThrGlyHisProValLeuThrCysGlnHisGlyThrAsnArgAsnTyrAspHis 1890
Db      1808 ATCTTGTAGGCCACCTGTGCTCACTGCGAGATGAACCTGACAAATCTGAACATAC 1867
Qy      1891 ProLeuProLysCysGlyValProCysGlyGlyAsnIleThrSerSerAsnGlyThrVal 1910
Db      1868 CTTTCCACAGGATGACGCTCCTGTGATATATGTGACATACGAATGTGACAT 1927
Qy      1911 TyrSerProGlyPheProSerProTyrSerSerSerGlnAspCysValTyrLeuIleThr 1930
Db      1928 TATTCCTCGGGTTTCAGATGAGTACCAATCTGAAGAGCTGTGTGGCTGTGACT 1987
Qy      1931 ValProIleGlyHisGlyValaGluLeuAsnLeuSerLeuLeuGlnThrGluProSerGly 1950
Db      1988 GTCCCTCAGAGCCAGAGGTATCAATTAATTAATCTGTGACGAGCCGAGGCTGTAAAT 2047
Qy      1951 AspPheIleThrIleTyrAspGlyProGlnGlnThrAlaProArgLeuGlyValaPheThr 1970
Db      2048 GACTACATTTGCTGTGTGATGTGCTGACCAAGAACTCGCTACGCTCGGGGTTTCACT 2107
Qy      1971 ArgSerMetAlaLysLysThrValGlnSerSerSerArgGlnValaLeuLeuLysPheHis 1990
Db      2108 GGAACACTGCGCTTGAAGACAGCTACAGCTCCACCAACAGAGGCTGTCAATTCAC 2167

```

|    |      |   |      |
|----|------|---|------|
| QY | 1991 | ArgAspAlaIaIaThnGlyGlyLeuPheAlaIaIaPheSerAlaTyPProLeuThnTlys    | 2010 |
| Db | 2168 | ACGGATTCTCCAAATGAGAGCTTCTTGTCTCCAAATTCATGCAATTCACCTGAAGG        | 2227 |
| QY | 2011 | CysPProPProThrlLeuPProAsnIaGluValIThnGluAsnGluIuPheAsn          | 2030 |
| Db | 2228 | TGCCGGCTCCCCCGACAGGTGCACAGGTGTACCTGTTTACAGAAAGIAGAGACTTTGAA     | 2287 |
| QY | 2031 | IleGlyAspIleValArgTyArgCysLeuPProGlyPheThnLeuValGlyAsnGluIle    | 2050 |
| Db | 2288 | AAAGGGAGCTTGTAAGTACAGAGCCATCCAGGGTACACATGTTGGGAAGGACACC         | 2347 |
| QY | 2051 | LeuThnCysIysLeuGlyThnTyTruEngInPheGluIysPProProIlaCysGluVal     | 2070 |
| Db | 2348 | CTGAATGCAAGCTCAAGCTCAACGTACTGTTCCAAAGCTCTCCACCACCTGTGAACA       | 2407 |
| QY | 2071 | HisCysProThrAsnGluLeuLeuThrAspSerThnGlyValIleLeuSerGlnSerTy     | 2090 |
| Db | 2408 | CAATGCCAGCAAAAGAGTGGAACAGAGTCTTGCGGGTGAATTCAGTCCGGGATC          | 2467 |
| QY | 2091 | ProGlySerTyPProGlnPheGlnThrCysSerTrpLeuValArgValGluPProAspTy    | 2110 |
| Db | 2468 | CCAGGCACATAATTTAACTCCAGACATGTGCTTGAGATTAATTAAGTGAACCAACTC       | 2527 |
| QY | 2111 | AsnIleSerLeuThnValGluTyPheLeuSerGluIysGlnTyAspGluPheGluIle      | 2130 |
| Db | 2528 | AACATCACACTTGTGTGGACACATTCCAAAGTAAAGCAATTTGATGACCTGGAAAGTG      | 2587 |
| QY | 2131 | PheAspGlyPProSerGlyGlnSerProLeuLeuValAlaLeuSerGlyAsnTySerAla    | 2150 |
| Db | 2588 | TTGATGTTCTTCTCGAACAAAGTCTCTGTGATGTCCTTAAGTGGGAACCAACGTGA        | 2647 |
| QY | 2151 | ProLeuIleValThrSerSerSerAsnSerValTyLeuAspGlyTrpSerSerAspHisIa   | 2170 |
| Db | 2648 | CAGTCCAAATTTTACAGACAGAAGTAAACCATCTGATCTCGCTGTCACAGACATGCA       | 2707 |
| QY | 2171 | TyrAsnAspGlyGlyPheIysIleArgTySerAlaProTyPProCysSerLeuProArgAla  | 2190 |
| Db | 2708 | ACCAGCAAGAAAGATTCACAGATTCGATNAGAGCTCTTACGACCTCACTCATCA          | 2767 |
| QY | 2191 | ProLeuHisIaGlyPheIleLeuGlyIlnThrSerThnGlnPProGlyIysIleHisPhe    | 2210 |
| Db | 2768 | CTCAGAGATGCGGCAATTTTAATATAAACCGGTGGGCGCATTTGGAGCAAGATCACTAC     | 2827 |
| QY | 2211 | GlyCysAsnAlaGlyTyArgLeuValGlyHisSerMetAlaIleCysThrArgHisPro     | 2230 |
| Db | 2828 | TTCTGCAAGCCTGGATATCGAATGATGTGGCCACAGCAATGGAACCTGCAGCGGAACCA     | 2887 |
| QY | 2231 | GlnGlyTyPThrIleLeuTrpSerGluAlaIleProLeuCysGlnAlaLeuSerCysGlyLeu | 2250 |
| Db | 2888 | GTGGGCGTGTAAACAGTGGGACTCGATGGACACACTTTCGAGCGTGTCTGTGGAAAT       | 2947 |
| QY | 2251 | ProGluAlaProTyAsnGlyMetValPheGlyIysGluTyPThnValGlyThyIysAla     | 2270 |
| Db | 2948 | CCAGAGGCTCCGGGAATGCGTCATTCACGGGCATATGATTCACCTTAGAAGATTAAGTG     | 3007 |
| QY | 2271 | ValTyPProCysSerGlyGlyTyPThrIleLeuGlnAlaGlyAlaThrAlaGluCys       | 2290 |
| Db | 3008 | ACTTATGATGATGAAGAGGCTTCAAGCTGATGCACTGACAGGAACCCACCACTGATGT      | 3067 |
| QY | 2291 | LeuAspThnGlyLeuTrpSerAsnArgAsnValPProProGlnCysValProValThnCys   | 2310 |
| Db | 3068 | CAGAAAGATGGCTTGATGAGACACAGAAAGCAACCCACATGCAACAGCTGCTGC          | 3127 |
| QY | 2311 | ProAspValSerSerIleSerValGlnHisGlyArgTrpArgLeuIlePheGluThnGln    | 2330 |
| Db | 3128 | CCAGACATCGAAGGCGACGTCGACAGACAGCGCTCGAGAGCTGATTTCGGATCACTG       | 3187 |
| QY | 2331 | TyrGlnPheGlnAlaGlnLeuMetLeuIleCysAspPProGlyTyPThrTyPThnGlyGln   | 2350 |
| Db | 3188 | AATGAATACGAGATCAAGATCTCTCTCACTGCTGCACGCCCGGCTACTCTTCCAGAGGTACG  | 3247 |

|    |      |  |      |
|----|------|--|------|
| QY | 2351 | ArgValIleArgCysGlnAlaAsnGlyLeuTrpSerLeuGlyAaspSerThrProThxCys      | 2370 |
| Db | 3248 | AGGCTCTTGACAGTGTCAAGCCAAAGGAGACCTGGAGCACTGAGAGACAGACCC-----        | 3301 |
| QY | 2371 | ArgIleIleSerCysGlyGluLeuProIleProProAsnGlyHisArgIleGlyThrLeu       | 2390 |
| Db | 3301 | -----  | 3301 |
| QY | 2391 | SerValTYrGlyAlaThrAlaIlePheSerCysAsnSerGlyTYrThrLeuValGlySer       | 2410 |
| Db | 3301 | -----  | 3301 |
| QY | 2411 | ArgValAlaGlyCysMetAlaAsnGlyLeuTrpSerGlySerGluValArgCysLeuAla       | 2430 |
| Db | 3302 | -----AGATGTAAGCG   | 3313 |
| QY | 2431 | GlyHisCysGlyThrProGluProIleValAsnGlyHisIleAsnGlyGluAsnTYrSer       | 2450 |
| Db | 3314 | GGTCATTGTGGCTCTCCAGACCCCATTTGTGAATGGCCATATCAAGCGGTGATGTTTCACG      | 3373 |
| QY | 2451 | TYrArgGlySerValTYrGlnCysAsnAlaGlyPheArgLeuIleGlyMetSerVal          | 2470 |
| Db | 3374 | TACAGAGACACAGTGGTCTACCAATGACCACTGGTTCGGGCTCGAGGGACACTGTGTG         | 3433 |
| QY | 2471 | ArgIleCysGlnGlnAspHisIleTrpSerGlyValSerThrProPheCysValProIleThr    | 2490 |
| Db | 3434 | AGAAATTGGCTCGAGGAGACCAACAGTGTGGGGGCACTCTGTTCGTCCTCCATACA           | 3493 |
| QY | 2491 | CysGlyHisProGlyAsnProValAsnGlyLeuThrGlnGlyAsnGlnPheAsnLeuAsn       | 2510 |
| Db | 3494 | TGTGGACACCTGGAAACCTCTGGCCATGGCTCAACAAGGCAACGAGTTCAACTTAAC          | 3553 |
| QY | 2511 | AspValValIlyPheValCysAsnProGlyTYrMetAlaGluGlyAlaAlaArgSerGln       | 2530 |
| Db | 3554 | GACCTTGAAATTTCACCTGGCAATACGTGGTACCGGCTCGAGGGGTGCTCCGAGCCAG         | 3613 |
| QY | 2531 | CysLeuAlaSerGlyGlnTrpSerAspMetLeuProThrCysArgIleIleAsnCysThr       | 2550 |
| Db | 3614 | TGTGGAGCAACGGCCAGTGGAGAGCCCTTGGCCATCTGTGACAGTGTAACTGTTC            | 3673 |
| QY | 2551 | AspProGlyHisGlnGlnIleAsnSerValArgGlnValHisAlaSerGlyProHisArgPhe    | 2570 |
| Db | 3674 | GACCCCGGATGTGGAAATATGCACTTGGCCATGGGCAACAGAACTTCCAGAGACTTC          | 3733 |
| QY | 2571 | SerPheGlyThrThrValSerTYrArgCysAsnHisGlyPheTYrLeuLeuGlyThrPro       | 2590 |
| Db | 3734 | GAGTATGGGACAAAGTGTGATGTATCACTGGAAAACGGGGTTTACTTGCTGGGACTTCT        | 3793 |
| QY | 2591 | ValLeuSerCysGlnGlyAspGlyThrThrAspArgProArgProGlnCysLeuLeuVal       | 2610 |
| Db | 3794 | GCCCTTACCTGCAGGCAAGTGGCTTGTGGAGCCGCTCTTACCACCAAGTCTGGGCTATA        | 3853 |
| QY | 2611 | SerCysGlyHisProGlySerProProHisSerGlnMetSerGlyAspSerTYrThrVal       | 2630 |
| Db | 3854 | TCAATGTGGCATCTCTGGAGTCCCTGCTAAATGCTGTCTGACTGGAGAAATGTTTACATAC      | 3913 |
| QY | 2631 | GlyAlaValAlaValArgTYrSerCysAlleGlyLeuArgThrLeuValGlyAsnSerThrArg   | 2650 |
| Db | 3914 | GGAAGCAACAGTTCACTCTCTGGAAGAGGGGGCCAGATCTCTCAAGGCAACAGACACAGA       | 3973 |
| QY | 2651 | MetCysGlyLeuAspArgGlyHisIleTrpThrGlySerLeuProHisCysSerGlyThrSerVal | 2670 |
| Db | 3974 | GTCGGCCAAAGACAGTCACTGGAGTGATCCCTTCTCATTTGTTCCAGGAAATATGCTCT        | 4033 |
| QY | 2671 | GlyValCysGlyAspProGlyIleProAlaHisGlyIleArgLeuGlyAspSerPheAsp       | 2690 |
| Db | 4034 | GGATTTTGTGTGATCCAGGAGACCCCAAGCAATGGGCTGTGCTTGGGGATGAGTTTAAG        | 4093 |
| QY | 2691 | ProGlyThrThrValMetArgPheSerCysGlyAlaIleGlyHisValLeuArgGlySerSerGln | 2710 |
| Db | 4094 | ACAAAGAGCTCTTTGGCTCTCTCTGTGGAATAGGGCCACCAAGCTCGGGGCTCTGCAGAA       | 4153 |
| QY | 2711 | ArgThrCysGlnAlaAsnGlySerTrpSerCysGlySerGlnProGluCysGlyValIleSer    | 2730 |

|    |      |  |      |
|----|------|--|------|
| Db | 4154 | CGCACAATGCCGTGTAATGATCTCTGGTCAGAGATCAAGCTCTGTGTGAGAGCCGTGCC        | 4213 |
| Qy | 2731 | CyeGlyAsnProGlyIYThrProSerAsnAlaArgValAlaPheSerAspGlyLeuValPhe     | 2750 |
| Db | 4214 | TGCGGAAACCCAGGACACGCCGACCAACGGATGATCTCCAGACAGTATGGATCCTTCC         | 4273 |
| Qy | 2751 | SerSerSerIleValIYTrgIuCyAArgIuGlyIYTrgAlaThrGlyLeuLeuSerArg        | 2770 |
| Db | 4274 | TCCAGCTCTGTCACTTAAGCTGTGGTGGAGAGGCTACAGAGCTGGGGCTATCATCTCGG        | 4333 |
| Qy | 2771 | HisCySerSerValaAngIYThrTrpThrGlySerAspProGluCySLeuValIleAsnCyS     | 2790 |
| Db | 4334 | CACGTGCACGGCAATGGGACGTGGACAGGACACAGCTCTGTGCTGACATATCATGCTGT        | 4393 |
| Qy | 2791 | GlyAspProGlyIleProAlaAsnGlyLeuArgLeuGlyAsnAspPheArgIYAsnIys        | 2810 |
| Db | 4394 | GGGATCTCTGGCACACCTGCCCAATGGCATTCAGATTGGGACAGACTTCATCTTCAACAG       | 4453 |
| Qy | 2811 | ThrValIYThrIYTrgIuCySValaProGlyIYTrpMetGluSerHisArgValaSerValLeu   | 2830 |
| Db | 4454 | ACCGTAGCTATCATGTCAGTCAACCCCGGCTACTTAATGGAGCCCAACATCAACCACTAC       | 4513 |
| Qy | 2831 | SerCyValThrIysAspArgIYThrTrpAsnGlyIYThrIysProValCySlyValaLeuMetCyS | 2850 |
| Db | 4514 | CGCTGCACCAAAAGATGGCACATGGATATAGAGCCGCCCTCTGAAAGCTGTCTGTGC          | 4573 |
| Qy | 2851 | LysProProProLeuIleProAsnGlyIYLeuValaGlySerAspPheMetTrpGlySer       | 2870 |
| Db | 4574 | AACACGCTCTCCAGTCCCAATGGAAATGGAGGGGTCAAGCTTCGATGGGGTCC              | 4633 |
| Qy | 2871 | SerValIYThrIYAlaCySLeuGluGlyIYTrgIuLeuSerIleuProAlaValaPheThrCyS   | 2890 |
| Db | 4634 | AGCATTAAGCTATCAGCTGTGTGAATGGCTACACAGCTCTCCACTGGCCATCTGTCTCGC       | 4693 |
| Qy | 2891 | GluIYIYAsnGlySerTrpThrGlyIuLeuProGluCySAspProValPheCyGlyIYAsp      | 2910 |
| Db | 4694 | GAAAGGGCTGGAGTGTGGAAAGAGAAATCTCTCAAGTCTTGC-GTCTTC-----             | 4733 |
| Qy | 2911 | ProGlyValaProSerArgIYArgIuAspArgIYPheSerIYArgSerSerVal             | 2930 |
| Db | 4743 | -----  | 4743 |
| Qy | 2931 | SerPheSerCyHisAspProLeuValaLeuValaGlySerProArgArgPheCyGlnSer       | 2950 |
| Db | 4744 | ---ATCCAGTGCAGAGCCCCCATGTGTAATGAGGATCTCTCCAGAGAACCTGCCAGGCC        | 4800 |
| Qy | 2951 | AspGlyIYThrTrpSerGlyIYThrGlnProSerCySLeuAspProThrLeuIYThrCyAla     | 2970 |
| Db | 4801 | GATGGGATATGGAATGGCATCCACGCCACTGTATAGNNNNNNNNNNNNNNNNNNNNNNNN       | 4860 |
| Qy | 2971 | AspProGlyValaProGlnPheGlyIleGlnIleAsnSerGlnIYTrgIuValaGlySer       | 2990 |
| Db | 4861 | NN         | 4920 |
| Qy | 2991 | ThrValaLeuPheArgCySLeuIYLeuGlyIYTrpLeuLeuGlnIYSerIYThrArgIYThrCyS  | 3010 |
| Db | 4921 | NN           | 4980 |
| Qy | 3011 | LeuProAsnLeuIYThrTrpSerGlyIYThrProProAspCySValaProHisHisCyAArgGln  | 3030 |
| Db | 4981 | NN           | 5040 |
| Qy | 3031 | ProGluIYThrProThrHisAlaAsnValaGlyValaLeuAspLeuProSerMetGlyIYTrpThr | 3050 |
| Db | 5041 | CCAGAGACCCACGGCATGCAGAGCTGTAGACCATCGATCTTCCAGCTTTGGCTACACT         | 5100 |
| Qy | 3051 | LeuIleIYThrProAlaArgArgAlaSerProSerArgValaIleProSerThrIYAlaProAl   | 3070 |
| Db | 5101 | TTAGTCTACACTGTGCACCCCGGATTTTCTCTGTGTGTGATCTGAACACAGAGACGTGT        | 5160 |
| Qy | 3070 | ArgArgMetAlaAlaGlyIuAlaIleSer                                      | 3079 |

| DB         | 5161   | AAAGCAGTATGGAATGACGACGAAACT | 5188        |
|------------|--|-----------------------------|-------------|
| RESULT 3   |  |                             |             |
| LOCUS      | AK081081   | 3882 bp                     | trna linear |
| DEFINITION | Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:1930082H09 product: CUB and Sushi multiple domains 1, full insert sequence.  |                             |             |
| ACCESSION  | AK081081   |                             |             |
| VERSION    | AK081081.1   | GI:26099665                 |             |
| KEYWORDS   | HTC; CAP trapper.  |                             |             |
| SOURCE     | Mus musculus (house mouse)   |                             |             |
| ORGANISM   | Mus musculus   |                             |             |
| REFERENCE  | Eukaryota; Neozoa; Chordata; Carnivora; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.   |                             |             |
| AUTHORS    | 1 Carninci, P. and Hayashizaki, Y.   |                             |             |
| TITLE      | High-efficiency full-length cDNA cloning   |                             |             |
| JOURNAL    | Meth. Enzymol. 303, 19-44 (1999)   |                             |             |
| MEDLINE    | 99279253   |                             |             |
| PUBMED     | 10349636   |                             |             |
| REFERENCE  | 2  |                             |             |
| AUTHORS    | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  |                             |             |
| TITLE      | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes   |                             |             |
| JOURNAL    | Genome Res. 10 (10), 1617-1630 (2000)  |                             |             |
| MEDLINE    | 20499374   |                             |             |
| PUBMED     | 11042159   |                             |             |
| REFERENCE  | 3  |                             |             |
| AUTHORS    | Shibata, K., Itoh, M., Aizawa, K., Nagaoke, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kaishogi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.   |                             |             |
| TITLE      | RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  |                             |             |
| JOURNAL    | Genome Res. 10 (11), 1757-1771 (2000)  |                             |             |
| MEDLINE    | 20530913   |                             |             |
| PUBMED     | 11076861   |                             |             |
| REFERENCE  | 4  |                             |             |
| AUTHORS    | The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.   |                             |             |
| TITLE      | Functional annotation of a full-length mouse cDNA collection   |                             |             |
| JOURNAL    | Nature 409, 685-690 (2001)   |                             |             |
| MEDLINE    |  |                             |             |
| PUBMED     |  |                             |             |
| REFERENCE  | 5  |                             |             |
| AUTHORS    | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.   |                             |             |
| TITLE      | Analysis of mouse transcriptome based on functional annotation of 60,770 full-length cDNAs   |                             |             |
| JOURNAL    | Nature 420, 563-573 (2002)   |                             |             |
| MEDLINE    | 6 (bases 1 to 3882)  |                             |             |
| PUBMED     |  |                             |             |
| REFERENCE  | 6 (bases 1 to 3882)  |                             |             |
| AUTHORS    | Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okamoto, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, K., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. |                             |             |
| TITLE      | Direct Submission  |                             |             |
| JOURNAL    | Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Sueno-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, url: http://genome-gsc.riken.jp/, Tel: 81-45-503-9222.   |                             |             |

## COMMENT

Ex:81-45-503-9216)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in Riken.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: <http://genome.gsc.riken.jp/>  
 URL: <http://fantom.gsc.riken.jp/>  
 Location/Qualifiers

## FEATURES

source

1..3882  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="PANTOM DB:B930082H09"  
 /db\_xref="taxon:10090"  
 /clone="B930082H09"  
 /issue\_type="cerebellum"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="10 days neonate"  
 233..3656  
 /note="CUB and Sushi multiple domains 1 (MGD|MG1:2137383,  
 GB|NM\_053171, evidence: BLASTN, 99%, match=3606)  
 putative"

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 3882  
 Score: 4805.50 Matches: 842  
 Percent Similarity: 82.54% Conservative: 179  
 Best Local Similarity: 68.07% Mismatches: 200  
 Query Match: 28.29% Indels: 16  
 Ds: 3 Gaps: 2

US-10-016-248-2 (1-3104) x AK081081 (1-3882)

QY 284 GYIIEPFOALATYRGIVARIGGLIYSEARGRPHENSHISGLIYASPTIRLEULYS 303  
 DB 2 GGCAATCCAGCCTTATGGAGAGGAGCTGGCAGCAGCTTCTTGACATGAGACAGCCTCACC 61  
 QY 304 PHEGLUCYSGINPROALAPHEGLULEUVALGILNYSALAIETHRCYSGINLYASAP 323  
 DB 62 TTGAGTGCCAGGAGCGTTGAGCTGTGGAGAGAGAGATTAATCTCCAGAAAAC 121  
 QY 324 ASNGINTPSEARALALYSLYSPROGLIYCYVALPHESERCYPHEPHEAPNETHRSER 343  
 DB 122 AACCAAGTGTGGGCAACAAGCCAGCTGTGTTCATGTTCTTCAACTCAGCGCA 181  
 QY 344 PROSERGLYVALLEUSERPROANLYRPROGLIYASPTIRGLIYANHSLEUHSICYS 363  
 DB 182 CCTCTGGATCATCCTCTGGCAAACTATCTGAGAGATGGAAGCAACAAGAACTGT 241  
 QY 364 VALTIRLEULILEUVALATGPROGLIYSEARGLIHSISLEUALAPHEASNPISLEASP 383  
 DB 242 GTGTGTGTATATATATGAGCTGGAGCGGAGTCACTTCAATGATTTGAT 301  
 QY 384 VALGLUPROGLINPHEAPPHLEUVALILEYASGLIYALATHRALGLUALAPROVAL 403  
 DB 302 GTGAACCTCACTTGAATCTTCCGCGCAAAAGATGAGATTTCTGACATCAACAGTC 361  
 QY 404 LEUGLIYTHRPHESEGLIYANGINLEUPROSERSEITLTHRSERSEGLIYHSIVALA 423  
 DB 362 CTGGGAACCTTCTCTGGCAATGAGTGGCTGCAACAGCTGCGCAAGTGGACATAGTA 421  
 QY 424 ARGLEUGLUPHEGLINTHRAPHISSEITRGLIYLYSARGIYPHEASNILETHRPHETHR 443  
 DB 422 CGCTGAGATTGATGATGATCACTCCACACTGGCAGAGTTTCAACATCAATATATAC 481  
 QY 444 THRPHEARGHISANGIUCYSPROAPPROGLIYVALPROVALASGLIYLYSARGPHEGLY 463  
 DB 482 ACATTTGTGCAGAACAGAGTCCACAGACCTTGGGATTTCTGTAAATGACGGGGTTTGA 541  
 QY 464 AAPPSEIRLEUGLILEUGLIYSEERISLEPHELEUCYASPSGLIYLYPHELEUGLIYTHR 483

DB 542 GACAGGTTTCTGCTGGGAATTCAGTGTCTCTCACTGATGATATGGCTTTGGAACT 601  
 QY 484 GINGLYSEIRGLIYHRIETHRCYVALLEULYSGIYSEVALIYTRPANSERALA 503  
 DB 602 CAGGTTTCTAGTCATACAGTGCATCTTGCAGAGAGAAAGTGTCTGGAGCTTACT 661  
 QY 504 VALLEUARGYSGIYUALAPROCYSGIYGLIYHSLEUTHSERPROSEGLIYTHRIELU 523  
 DB 662 GTCCCTGCTGTGAACCTCATGTGTGGCACTGACCGCTTCCAGTGGGCTATATTA 721  
 QY 524 SERPROGLIYTRPROGLIYPHETRYLYSAPALALEUSERCYAALATRYVALILEGLUAL 543  
 DB 722 CCTCCAGATGGCGGGGATTAACAAGATTTCTTAACTGTAATGGGTCAATGAAGA 781  
 QY 544 GINPROGLIYTRPROILEYSLIETHRPHEAPAPGPHELYTHGLIYVALIENYTRASP 563  
 DB 782 AAMCAGACATTTCCATCAAGATTAACATTCAGACAGTTCCAGACGAAGTCAATATGAT 841  
 QY 564 THIRLEUGLIVALARGASPIYARGTHRYSERALAPROLEULILEGLIYVALIYRHSGLY 583  
 DB 842 ACCCTGGAAATACGGGATGGCCACACAGCTCATCCCACTCATAGAGATACACGGC 901  
 QY 584 THIRGLIYVALPROGLINPHEULIETSERTHSERANTYRLEUTYRLEULPHESETHR 603  
 DB 902 ACCCAGGACCCACAGATCTCTATCAAGACAGGACCACTACATGTAACCTGTTTACCACT 961  
 QY 604 AAPPYSEIRHSISERASPIIISGLIYPHEGLINLEUARGIYRGLIUTRILETHLEUGLISER 623  
 DB 962 GATACACAGCCGGGCTAGAGTGGCTTCTCTCATTCACATAGAGATGATCACTTGAATCC 1021  
 QY 624 ASPHSISCYLEUASPPROGLIYILEPROVALISNGIYGLIARGHISGLIYASNPSPHETRY 643  
 DB 1022 GACTCTGTGTGAGACCCAGGATCCCTGTAAATGTGACGCCACATGACATTAACCTTGGT 1081  
 QY 644 VALGLIYALALEUVALIETHRPHESECYASPOSEIRGLIYTHRIEUSERASPIYGLIYPRO 663  
 DB 1082 ATCAGGTCTAAGTACACTTCACTGAGCTGTGACCTGGATACACTCACTGAGATGAGTACCCC 1141  
 QY 664 LEUGLIYCYSGIUPROASNPHENGLINTPSEARGLIALAUEPPOSECYSGIUALALEUCYS 683  
 DB 1142 CTCTCTGTGTGAAGAACCATCAATGAGCAACGCTTGGCCAGCTGTGATGCTGTGT 1201  
 QY 684 GLYGLYPHEIIEGLINLYSEIRSEGLIYTHRIILEUSERPROGLIYPHEPPOAPPHETRY 703  
 DB 1202 GAGGCTTACATCTATGAAAGAGTGGAGCTGCTTCTCAGAGATTTCCAGACTTTAT 1261  
 QY 704 PROASNANLEUANCYETHRTPILEILEGLIUTHSERHSISGLIYSGIYVALIPHEPHE 723  
 DB 1262 CCAAACTCTGAACTGATACATGAGCAATGAGTCTCTATGAGCAAGGAGTGCAGATG 1321  
 QY 724 THRPHEIHSIETHRPHEHSISLEUGLUSERGLIYHSAPTRYLEULIETHRGLIUAENGILY 743  
 DB 1322 AATTTCAACACTTTTCACTTGAAGTCCCAAGACTATTTGTGTATTCAGAGATGGG 1381  
 QY 744 SEEPHETHRGLINPROLEUARGINLEUTHRGLIYSEARGLIYSEARGLIYPROLABROLLESERALA 763  
 DB 1382 AGTTTCTCAAGCAGCATAGCCAGGCTCATGGGTCACTGCTTCAACCAATTAAGGCT 1441  
 QY 764 GLIYLEUTYRGLIYANPHEHRIALAGIYVALIARGPHEIISERASPPHESEMERSETRYR 783  
 DB 1442 GGTTTGTGGAAACTTCACTGCAACACTCAAGTTTCACTGCACTTCCATCTTCCTAT 1501  
 QY 784 GLIUGIYPHEASNILETHRPHESEIRGLIYTRAPPLEUGLUPROCYSGIYGLIYVALI 803  
 DB 1502 GAAGGCTTCAACATTAAGTTTGCAGAAATGACCTGGAGCCCTGTGATGATCACTCGAGTC 1561  
 QY 804 PROALITRSEIRILEARGIYSGIYLEUGINPHEGLIYVALIYASPTIRLEUTHRPHESEIR 823  
 DB 1562 CTGCGCTTCAAGCCGAGATTTGGGTTTCAAGTTGAGTGTGGGAGAACCTTGGCTTTACCC 1621  
 QY 824 CYEPHEPPOGLIYTRARGLEUGLILYTHRALARGIETHRCYALEUGLIYGLIYARGARG 843

Db 1622 TGCTTCCAGGATACCGCTAGAGAGTGCACCAAGCTTACCTCGCTGGTGGGAGC 1681  
 Oy ArgLeuTPSerSerProLeuProArgCyValAlaGluCyGluYAsnSerValThrcly 863  
 Db 1662 CGAGGTGAGAGGACCTCTGCACAGGTGTGTGGCTGAATGGAGCAAGGCTCAAGGA 1741  
 Oy ThrcInglYThrLeuLeuSerProAsnPhcProValaenTYrAsnAsnAsnHleGluCyS 883  
 Db 1742 AATAGAGACATTAATCTCTCCAAATTTCCATCCATTAATGATTAATACATTAAGT 1801  
 Oy HleTYrSerHleGlnThrcInpProGlyYAsnGlyHleGlnLeuValaArgAlaHcGlu 903  
 Db 1802 ATCTATAAAATAGAAACAGAACCGGAGGAGATCCATCTCAGAGCTCGAACTTCCAG 1861  
 Oy LeuSerGluYAsnValLeuYValaYTrAspGlyAsnAsnAsnSerAlaArgLeu 923  
 Db 1862 CTCTTCGAGGAGACCTTAAGGTTTACATGAGAAAGACAGCTCTCAAGGCTCACTG 1921  
 Oy GlyValPheSerHleSerGluMerGlyValThrcLeuAsnSerThrcSerSerLeu 943  
 Db 1922 GGTGCTTACAGAGAGGAAATTCATGGGGCTGTGGCTCAACAGACCTCCACACACTG 1981  
 Oy TrpLeuAspPheHleThrcAspAlaGluAsnThrcSerGlyPheGluLeuHlePheSer 963  
 Db 1982 AGGCTGAGTTCACACCAACGAGGTGAGACACCGCCCAAGGCTTCACACTCACTACAC 2041  
 Oy SerPheGluLeuHleYAsnGluAspProGlyThrcProYAspPheGlyTYrLeuValHle 983  
 Db 2042 AGTTTACCTAGTAAATGTAGAGATCAGGACATCCCTAATAGCTTACAGGATCCGT 2101  
 Oy AspGluGlyHleAspHleAlaGlySerSerValSerPheSerCyAspProGlyTYrSerLeu 1003  
 Db 2102 GATATGGCCACTTACAGACACTAGTCTCTACAGCTGCAACCGGGCTACGCAATG 2161  
 Oy ArgGlySerGluGluLeuLeuCyLeuSerGlyGluArgArgThrcTrpAspArgProLeu 1023  
 Db 2162 CATGGCAGACATACCTGACCTGCTTGTAGTGGGAGCCGAGGCTGTGGAGCAAACTATG 2221  
 Oy ProThrcYValAlaGluCyGlyGlyThrcValaArgGlyGluValSerGlyGluValLeu 1043  
 Db 2222 CTTTCCTGTGTGGCTGAATGTGTGTCTTGTTCATGACGACACATCGGGGCGCTACTG 2281  
 Oy SerProGlyTYrProAlaProTYrGlyHleAsnLeuAsnCyHleTrpThrcHleGluAla 1063  
 Db 2282 TCTCTGGCTACCCGCCCCATATACAAACACTTCTCACTGACCTGACCATAGAGCC 2341  
 Oy GluAlaGlyCyThrcHleGlyLeuHleAspHleValaPheAspThrcGluValHleAsp 1083  
 Db 2342 GATCCTGGAAGACATCAGCTCATTTTCAATTTGTGACACTGAAACGGCGCATAC 2401  
 Oy ValLeuAlaGlyETrPAspGlyProValaGluSerGlyValLeuLeuYAsnSerGly 1103  
 Db 2402 ATCTTCAGAGTCTGGAGTGGCTGTAGACAGCAATCTCTGAGAGAGTGGAGCGCT 2461  
 Oy ProAlaLeuProYAspLeuHleAsnSerThrcPheAsnSerValaLeuGluPheSerThrc 1123  
 Db 2462 TCGGCTCACTGAGAGATCAGACGACCTTCACTCACTCACTCACTGAGTTGACAGT 2521  
 Oy AspPhePheThrcSerYAsnGlyPheAlaHleGlnPheSerValSerThrcAlaThrcSer 1143  
 Db 2522 GACTTCTTCATCAGAAAGTCCGGCTTCCATCCAGTTTCCACTTCATTCAGATCCACC 2581  
 Oy CyAsnAspProGlyYAsnGlyAsnGlySerArgSerGlyYAspSerTrpGluAlaGly 1163  
 Db 2582 TGCATATGACCTGGAGATCCTCAGATGTATCCCTCGGTGACAGCGGAGAACTTGA 2641  
 Oy AspSerThrcValPheGlnCyAspProGlyTYrAlaLeuGlnYAsnAlaGluHleSer 1183  
 Db 2642 GACACCATACCTTCAAGTGAACCTTGATACAGCTTCAAGAGCCAGCCAGATATCAG 2701  
 Oy CyValaYAsnGlyAsnArgPhePheThrcInpProSerProProThrcCyAlaLeuAlaPro 1203  
 Db 2702 TGTGTGACAGTGAACACGTTTCTTGTGGAGCCAGATCTCTCTGTGATAGCTCT 2761

Oy 1204 CyGlyGlyAspLeuThrcGlyProSerGlyValaHleLeuSerProAsnTYrProGluPro 1223  
 Db 2762 TGTGTGAGAAATCAGACGCTCTGTGAGATGATCTATCCCAATACCAACAGCA 2821  
 Oy TYr-ProProGlyYAsnGlyCyAspTrpYValaThrcValSerProAspTYrValaHle 1243  
 Db 2822 TACCTCTCTGAGAGAGTGCAGCTGGAGGATTAAGGTGAACCAAGCACTTGTGATTC 2881  
 Oy AlaValaPheAsnHlePheAsnLeuGluProGlyTYrAspPheLeuHleGlyYAspGly 1263  
 Db 2882 CTGATATTCAAAAGTTTACATGAGAGCCAGATTATGATCTCTGATATCTATGAGG 2941  
 Oy YArgAspSerLeuSerProLeuHleGlySerPheTYrGlySerGluLeuProGlyYArg 1283  
 Db 2942 GAGAGACTCAACGCCCATGATTTGAGAGTTTCCAGGCTTCAAGCCCAAGAGGAT 3001  
 Oy eGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerValSerAsnAl 1303  
 Db 3002 CGAGAGCAGTGTAAACAGCTCTTCTGGCATTCAGAGATGATCTTGTGGCCGTGTC 3061  
 Oy aGlyPheValaHleAspTYrThrcGluAsnProArgGluSerCyAspAspProGlySer 1323  
 Db 3062 GGGATTCGCCATTTGAATTAAGAGAAACCCCGGAGAGCTTGTGACCTGGGAGACAT 3121  
 Oy eLYAsnGlyYThrcArgValaGlySerAspLeuYLeuGlySerSerValaThrcTYrCY 1343  
 Db 3122 AATGAATGGACAAAGATGTGAACTGATTTAACTGGGCTTACAGTACTTATCACTG 3181  
 Oy SHleGlyYTYrGluValaGluGlyThrcSerThrcLeuSerCyAlaLeuGlyProAspGly 1363  
 Db 3182 TGACTGTGTAAACAGATGTGATGCCCTGCTCATTTGAGTGTGACAGGGGCTGATG 3241  
 Oy YLYAspProValaTrpAsnAsnProArgProValaCyThrcAlaProCyGlyGlyGlyTYr 1383  
 Db 3242 GAACCGCTCCGAGGAGCGGACATCACTGCTGCAAGCAACCTGTGGAGGCCAGTTCAC 3301  
 Oy IGLYSerAspGlyValaLeuSerProAsnTYrProGluAsnTYrThrcSerGlyGln 1403  
 Db 3302 GGGCTCGAGGGGTAGTGTGTGACCAAACTACCTTCATATACACAGCTGGAGCAT 3361  
 Oy eCyLeuTYrPheValaThrcValaProYAspTYrValaPheGlyGlnPheAlaPhePh 1423  
 Db 3362 GTGATATCTATTCATCAGGCGCCCAAGAAATGTGTGTATTTGACAGTTGCTATTT 3421  
 Oy eHleThrcAlaLeuAsnAspValaGluValaHleAspGlyHleSerGlnHleSerArgLe 1443  
 Db 3422 CCAGACTGCACTGAACGACTTGGCAGAGTTGTGTGATGAAACACACACCCCGAGCGCT 3481  
 Oy uLeuSerSerLeuSerGlySerHleThrcGlyGluSerLeuProLeuAlaThrcSerArg 1463  
 Db 3482 TCTGAGTCTCTCTGTGTGTCTCCATTCAGTGAAGAACTGCACTGGCTATATCAATCA 3541  
 Oy nValLeuHleYAspPheSerAlaYAsnGlyLeuAlaProAlaArgGlyPheHlePheVala 1483  
 Db 3542 GATTCGTCTGATTCAGCGCCCAAAAGCGAGCTTCTGACGGGTTTCCACTTCGTCTA 3601  
 Oy rGlnCySerSerSerValProGluProArgTYrGlyYAsnArgLeuGlySer 1508  
 Db 3662 ATATTGT-----GTACCAAGAGAGAGGTGCTCCCGGCTCC 3698

RESULT 4  
 LOCUS AY407074 5824 bp DNA linear GSS 15-DEC-2003  
 DEFINITION Pan troglodytes CSMD1 gene, VIRUAL TRANSCRIPT, partial sequence,  
 genomic survey sequence.  
 ACCESSION AY407074  
 VERSION AY407074.1 GI:39763045  
 KEYWORDS GSS.

SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes  
 REFERENCE Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Eukaryota: Eutheria: Primates: Catarrhini: Hominoidea: Pan.  
 AUTHORS 1 (bases 1 to 5824)  
 TITLE Clark, A.G., Glanowski, S., Nelson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civeille, D.R., Lu, F., Murphy, B., Ferrieres, S., Wang, G., Zheng, X.H., White, T.J., Shtinsky, J.J., Adams, M.D. and Cargill, M.  
 JOURNAL Science 302 (5652), 1960-1963 (2003)  
 PUBMED 14671302  
 REFERENCE 2 (bases 1 to 5824)  
 AUTHORS Clark, A.G., Glanowski, S., Nelson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civeille, D.R., Lu, F., Murphy, B., Ferrieres, S., Wang, G., Zheng, X.H., White, T.J., Shtinsky, J.J., Adams, M.D. and Cargill, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA  
 COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.  
 FEATURES  
 source location/Qualifiers  
 1..5824  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 <1..>5824  
 /gene="CSMD1"  
 /locus\_tag="HCM2769"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 5824  
 Score: 4001.00 Matches: 720  
 Percent Similarity: 63.68% Conservative: 176  
 Best Local Similarity: 51.17% Mismatches: 453  
 Query Match: 23.56% Indels: 58  
 DB: 9 Gaps: 1  
 US-10-016-248-2 (1-3104) x AY407074 (1-5824)  
 QY 1320 ProGlySerIleuYsaenglyThaYvaGlySerlepleuYleuglySerSerVal 1339  
 DB 2 CCAGAAATATTAATGAATGGAGCAAGAAATGGAAACAGACTTCAAGCTTGCTCCACATC 61  
 QY 1340 ThTYTYTCYshIsGlyGlyTYrGluValGluGlyThSerThreusSerCysIleleu 1359  
 DB 62 ACCTAACAGTGTACTCTGCTCAACAAGATCTTGACCCCTCAATCCATCACTGTGTGATT 121  
 QY 1360 GlyProAspGlyLySP-roValITTPAsnAsnProArgProValCysThrAlaProCysGly 1379  
 DB 122 GGGGCCCATGGGAAACCTCTGCGGACCAAGTCTGCTCCCTCGCAATGCTCCCTGTGA 181  
 QY 1380 GlyGlnTYrValGlySerAspGlyValValLeuSerProAsnTYrProGlnAsnTYrThr 1399  
 DB 182 GGCACAGTACACGGGATAGAGGGGTATTTATCAACCAATCAACCCCATATTAACACA 241  
 QY 1400 SerGlyGlnIleCysleuTYrPheValThValProLYAspTYrValValPheGlyGln 1419  
 DB 242 GCTGCTCAATATGCTCTATTCATCAAGTCAAGTTCGCAAGAAATTCGNNNNNNNNNN 301  
 QY 1420 PheAlaPhePheHisThrAlaLeuAsnAspValValGluValHisAspGlyHisSerGln 1439  
 DB 302 NNNNNNNNNNNNNNNNNNACAGCCCTGAATGATTGGCGGAATTTATGATGAACCATGCA 361  
 QY 1440 HisSerArgleuLeuSerSerleuSerGlySerHisThGlyGlnSerleuProleuAla 1459  
 DB 362 CAGGCGACACTTCTACGCTGCTCTGCGGCTCTCACTCAAGGGAACATGCTTGCTT 421  
 QY 1460 ThSerAsnGlnValleuIleYspPheSerAlaYsglyLeuAlaProAlaArgGlyPhe 1479

DB 422 AGCTCAATCAAAATTCGTCCGATTCAGTCAAGAGCGGTCTCTGCCCGGGCTTC 481  
 QY 1480 HisPheValTYrGlnAlaValProArgThnSerAlaThGlnCysSerSerValProGlu 1499  
 DB 482 CACTTCGTGTATTAAGCTGTTCTCTGACAGGACCAACCAAGAGCTTGTGCCCGAG 541  
 QY 1500 ProArgTYrGlyYsaArgleuGlySerAspPheSerValGlyAlaIleValArgPheGlu 1519  
 DB 542 CCAGATACCGAAGAGAAATGGTCTGAGTTTGGCGGGCTCCATCTCCGATTCCAG 601  
 QY 1520 CysAsnSerGlyTYrValleuGlnGlySerProGlnIleGluCysleuProValProGly 1539  
 DB 602 TGCACCCGGGATACCTGCTCAAGGCTTCACGGGCTCCACTGCCAGTCCGCTCCAT 661  
 QY 1540 AlaLeuAlaGlnTPAsnValSerAlaProThrCysValValProCysGlyGlyAsnleu 1559  
 DB 662 GCCTTGACAGAGTGAAGACAGACAGATCCCGACTGTGTGTCCTGCACTGGCAATTC 721  
 QY 1560 ThGluArgArgGlyThrIleleuSerProGlyPheProGluProTYrleuAsnSerleu 1579  
 DB 722 ACTCAAGAGAGGATCAATCTCTCCCGCTACCTGAGCCATACGGAACCACTTG 781  
 QY 1580 AsnCysValTrpIleValIleValProGluGlyAlaGlyIleGlnIleGlnValSer 1599  
 DB 782 AACTGTATATGAAGATCACTACTTACGAGGCTCGGANNTTCAATCCAGTCAATCACT 841  
 QY 1600 PheValThrGluGlnAsnTPAspSerleuGluValPheAspGlyAlaAspAsnThrVal 1619  
 DB 842 TTTGCCACGAGACAGAACTGGGACTCCTTGAGATCCAGATGGTGGAGATGTACCCCA 901  
 QY 1620 ThrMetleuGlySerPheSerGlyThrThrValProAlaLeuAsnSerThSerAsn 1639  
 DB 902 CCGAGACTGGAGAGCTTCTCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 961  
 QY 1640 GlnleuTYrleuHisPheTYrSerAspIleSerValSerAlaIleGluPheIleGlu 1659  
 DB 962 NNN 1021  
 QY 1660 TYrIleThrValGlyleuSerSerCysProGluProAlaValProSerAsnGlyValYs 1679  
 DB 1022 NNNNNNNCTGTAGGCTTGTGTCGATGCAAGAACACACCCCTCCCAACACAGATCAA 1081  
 QY 1680 ThrGlyGluArgTYrleuValAsnAspValValSerPheGlnCysGluProGlyTYrAla 1699  
 DB 1082 ATCGAGATTCGATACATGATGACAGATGCTCTCTTCCAGTGGAGCCCGGGTAAACC 1141  
 QY 1700 LeuGlnGlyHisAlaHisIleSerCysMetProGlyThrValArgAGTTPAsnTYrPro 1719  
 DB 1142 CTGACAGGCGCGTCCCATTTCTGTATGCGAGGACCGTGGCGGTGGAATATCCG 1201  
 QY 1720 ProProleuCysIleAlaGlnCysGlyGlyThrValGluGluMetGluGlyValIleleu 1739  
 DB 1202 TCTCCCTGATGATGCACTGAGGAGGAGCTGAGCACTTGGGTGTGTGATCTCTG 1261  
 QY 1740 SerProGlyPheProGlyAsnTYrProSerAsnMetAspCysSerTPAsnIleAlaIleu 1759  
 DB 1262 AGCCCGGCTTCCGGGTTCTTACCCCAACATTAAGTCACTGAGAGATCTCACTTA 1321  
 QY 1760 ProValGlyPheGlyAlaHisIleGlnPheleuAsnPheSerThGluProAsnHisAsp 1779  
 DB 1322 CCCATCGCATAGTGCACATATTCAGTTTCTGAATTTTCTACCAAGCTATATCATGAC 1381  
 QY 1780 TYrIleGluIleArgAsnGlyProTYrGluThSerArgMetMetGlyArgPheSerGly 1799  
 DB 1382 TTCTTAAATTAATTAATGAACCTTACACACAGGCCCATGATGAGCAATTTTACCGCC 1441  
 QY 1800 SerGluLeuProSerSerleuLeuSerThSerHisGluThrThrValTYrPheHisSer 1819  
 DB 1442 ACGATCTCCCGCGGCGCTGACACAGCAAGCAATGATTCATCACTTTATAT 1501  
 QY 1820 AsnHisSerGlnAsnArgProGlyPheYsleuGluTYrGlnAlaTYrGluLeuGlnGlu 1839  
 DB 1502 GACCAATTCGCAAAACCGGAGAGATTAACTTGCTTACCAAGCTTATGATTAACAGAAC 1561

1840 CybProAspProGluProPheAlaAsnGlyTLeValArgGlyAlaGlyTyrAsnValGly 1859  
1562 TGTCAATGATCCACCCCATTTTCAGAAATGGGATCATGATCAATGATTAAGGGTGGG 1621  
1860 GlnSerValThrPheGluCysLeuProGlyTyrGlnLeuThrGlyHisProValLeuThr 1879  
1622 CAATCAATATCTTCCAGGTGTATCTGGGTACATTTCTAATAGCCATCTGCTCTACT 1681  
1880 CysGlnHisGlyThrAsnArgAsnTyrAspHisProLeuProDlyGlyValProCys 1899  
1682 TGTGACGACGGGATCAACAGAACTGGAATCACTTTTCCAAATGATGAAGCCCTGTG 1741  
1900 GlyGlyValLeuIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyr 1919  
1742 GGGTACAACTGATCTTCTCAGAACGGCACATCTACTCTCCGTGCTTCTCATGATGAT 1801  
1920 SerSerSerGlnAspCysValTyrPheLeuIleThrValProIleGlyHisGlyValArgLeu 1939  
1802 CCGATCCGTAAGAGCTGATTTGGCTCATCATCAGTGCTCCGGGGCAGGAGTTTACATC 1861  
1940 AsnLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTyrAspGlyPro 1959  
1862 AACTTCACTCTGTACAGACGGAGCTGTCAACGATTAATGCTGTGTTGNNNGTCCC 1921  
1960 GlnGlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaValGlyThrValGln 1979  
1922 GATCAGAACTCAACCCAGCTGGAGTTTTCAGTGCAACACAGCTCTGNNNNNNNNAT 1961  
1980 SerSerSerAsnGlnValLeuLeuValPheHisArgAspAlaAlaThrGlyIlePhe 1999  
1982 AGCTCCACCAACCAAGTCTGCTCAAGTTCACAGCCAGCTTTCAATGAGAGCTTCTTT 2041  
2000 AlaIleAlaPheSerAlaTyrProLeuThrLysCysProProThrIleLeuProAsn 2019  
2042 GTCTCTCAATTTCCAGNN 2101  
2020 AlaGlyValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrArgCys 2039  
2102 NNN 2161  
2040 LeuProGlyPheThrLeuValGlyAsnGlyIleLeuThrCysValLeuGlyThrTyrLeu 2059  
2162 CACCCGGGTACACTTGGTGGGACGACACTNNAGCTTCCAGCTCAGTCCAGTTG 2221  
2060 GlnPheGluGlyProProIleCysGlyValHisCysProThrAsnGluLeuLeuThr 2079  
2222 CAGTTGAGGGTCCCTCCCAACATGTGAAGNNNNNNNNNNNNNNNNNNNNNNNNNN 2281  
2080 AspSerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThr 2099  
2282 NNN 2341  
2100 CysSerTyrLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPhe 2119  
2342 NNN 2401  
2120 LeuSerGlyGlyTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerPro 2139  
2402 NNN 2461  
2140 LeuLeuValAlaLeuSerGlyValAsnTyrSerAlaProLeuIleValThrSerSerSerAsn 2159  
2462 CTGTGTGTAATCTTAAGTGGAAATCATGAAATCAATCAATTTTACAGAGGAGTAT 2521  
2160 SerValTyrLeuArgTyrPheSerSerAspHisAlaTyrAsnArgGlyGlyPheValArg 2179  
2522 CAGTTAATCTCCGCTGCTCAGTCACTGACCATCCACAGTAAAGAAATTTCAATTCGC 2581  
2180 TyrSerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGln 2199  
2582 TATGACACACTTACTGACATTTGACCCACCCGCTGAAGAAATGGGTAATTTCAACAGA 2641

2200 ThrSerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuVal 2219  
2642 ACTCAGAGACCGGTGGAGAACCAAGTCAATATTTTTCAGAGCTGATACCAATGCTC 2701  
2220 GlyHisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuTyrPheSerGluAla 2239  
2702 GGCCACAGCAATGCACTGTAGACGAACCACTTGTCATGTACAGTGGGACTCCCTC 2761  
2240 IleProLeuGlySerGlnAlaLeuSerCysGlyLeuProGluAlaProDlyAsnGlyMetVal 2259  
2762 ATGCACCTTGGCCAGGCTGTGTCTGTGGAATCCCAAAATCCCGAGAAACGGTTCAATT 2821  
2260 PheGlyLysGlyTyrThrValAlaGlyThrValAlaValTyrSerCysSerGlnGlyTyrHis 2279  
2822 ACCGGGAACAGATTCCTTTGGACAGTAAGTGTCTATGAAATCTCATAGAGCTTCAAG 2881  
2280 LeuGlnAlaGlyAlaGluAlaThrAlaGlyCysLeuAspThrGlyLeuTyrPheSerAsnArg 2299  
2882 CTGAAATCCAGCCAGCAAGCAACGCGTGTCTCAAGAAAGATGGGCTGTGGATGACAG 2941  
2300 AsnValProProGlnCysValProValThrCysProAspValSerSerIleSerValGlu 2319  
2942 GGGAAAGCCGCCACGTGTAAAGCCGCTCTTGCCTCCGACATGAAAGCTCAGCTTCAGAA 3001  
2320 HisGlyArgTyrPheGlyLeuIlePheGluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeu 2339  
3002 CATCTCATCTGGAGAGCTGCTTTCAGATCTCTGAAATAGTACGGTCTCAAGTATGCTG 3061  
2340 IleCysAspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGly 2359  
3062 AGCTGACAGCTCGTGTATCTACTTGAAGGCGGAGGCTCCTCGACGTCTCAGGCCAATGGG 3121  
2360 LysTyrSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeuPro 2379  
3122 ACGTGAACGTAGAGATGAGAGGCGCAAGCTGTGA----- 3157  
2380 IleProProAsnGlyHisArgIleGlyThrLeuSerValTyrGlyAlaThrAlaIlePhe 2399  
3157 ----- 3157  
2400 SerCysAsnSerGlyTyrThrLeuValGlySerArgValArgGluCysMetAlaAsnGly 2419  
3157 ----- 3157  
2420 LeuTyrPheSerGlySerGluValArgCysLeuAlaGlyHisCysGlyThrProGluProIle 2439  
3158 -----GCTGGCCACTGCGGCTTCCCAAGCCCATTT 3187  
2440 ValAsnGlyHisAlaAsnGlyGluAsnTyrSerTyrArgGlySerValIleTyrGlnCys 2459  
3188 GTGAACGCTCACTTACTGTGAAGATGGCTTCAAGTACAGACAGGAGTGTATACCAAGTGC 3247  
2460 AsnAlaGlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisIleTyr 2479  
3248 AATCTGTGTTCCGGCTGTGGGAATCTCGGTAGATATGCTTCAAAAGCAACAAGTGG 3307  
2480 SerGlyLysThrProPheCysValProIleThrCysGlyHisProGlyAsnProValAsn 2499  
3308 TCTGACAAACCGCTGTGTGTGCTCCCATCACTGTGTGTCACCTCGAAGAACCCGCCAC 3367  
2500 GlyLeuThrGlnGlyAsnGlnPheAsnLeuAspValValLysPheValCysAsnPro 2519  
3368 GGAATTCATTAATGGACAGTGAAGTTCAACTTAATGACATCGTGAATTCACCTGACACAG 3427  
2520 GlyTyrMetAlaGluGlyAlaAlaArgSerGlnCysLeuAlaSerGlyGlnTyrPheSerAsp 2539  
3428 GGGATATTTGCGAGGGCGGTGTCTCGAGCCAGTGTGAGCAACGCGCACTGGAGTACG 3487  
2540 MetLeuProThrCysValGlyIleIleAsnCysThrAspProGlyHisGlnGluAsnSerVal 2559  
3488 CTTCTGCCACAGTGTCAAGTGTGAAGTCTTATCATCAGAGCTTTTNGTAAATGNNNNN 3547  
2560 ArgGlnAlaHisAlaSerGlyProHisArgPheSerPheGlyThrThrValSerTyrArg 2579



|            |   |  |                             |
|------------|---|--|-----------------------------|
| Db         | 660   | AACGCGCTTACTCAGGAGGACCAAGTTCACCTCAACCTCAGACGTCGTCAAGTTCATTTCGANC | 713                         |
| Qy         | 2519  | Progl1YrMecAlaGluc1YAlaAla                                       | 2527                        |
| Db         | 720   | CNAGGATACATGACTGAGGGGCTCG  | 746                         |
| RESULT 6   |   |  |                             |
| LOCUS      | AK046385  | 2841 bp  | mRNA linear HTC 03-APR-2004 |
| DEFINITION | Mus musculus adult male copro quadrigemina cDNA, RIKEN full-length enriched library, clone:2203078117 product:hypothetical Sushi domain / SCR repeat / CDP module/ CUB domain containing protein, full insert sequence.   |  |                             |
| ACCESSION  | AK046385  |  |                             |
| VERSION    | AK046385.1  | GI:26091501  |                             |
| KEYWORDS   | HTC, CAP trapper.   |  |                             |
| SOURCE     | Mus musculus (house mouse)  |  |                             |
| ORGANISM   | Mus musculus  |  |                             |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  |  |                             |
| AUTHORS    | Carninci, P. and Hayashizaki, Y.  |  |                             |
| TITLE      | High-efficiency full-length cDNA cloning  |  |                             |
| JOURNAL    | Meth. Enzymol. 303, 19-44 (1999)  |  |                             |
| MEDLINE    | 99279253  |  |                             |
| PUBMED     | 10349636  |  |                             |
| REFERENCE  | 2   |  |                             |
| AUTHORS    | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.   |  |                             |
| TITLE      | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  |  |                             |
| JOURNAL    | Genome Res. 10 (10), 1617-1630 (2000)   |  |                             |
| MEDLINE    | 20499374  |  |                             |
| PUBMED     | 11042159  |  |                             |
| REFERENCE  | 3   |  |                             |
| AUTHORS    | Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.   |  |                             |
| TITLE      | RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer   |  |                             |
| JOURNAL    | Genome Res. 10 (11), 1757-1771 (2000)   |  |                             |
| MEDLINE    | 20530913  |  |                             |
| PUBMED     | 11076861  |  |                             |
| REFERENCE  | 4   |  |                             |
| AUTHORS    | The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  |  |                             |
| TITLE      | Functional annotation of a full-length mouse cDNA collection  |  |                             |
| JOURNAL    | Nature 409, 685-690 (2001)  |  |                             |
| MEDLINE    | 5   |  |                             |
| PUBMED     |   |  |                             |
| REFERENCE  |   |  |                             |
| AUTHORS    |   |  |                             |
| TITLE      | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  |  |                             |
| JOURNAL    | Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  |  |                             |
| MEDLINE    | Nature 420, 563-573 (2002)  |  |                             |
| PUBMED     | 6 (bases 1 to 2841)   |  |                             |
| REFERENCE  | Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hizomoto, K., Hirooka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, K., Ohno, M., Ohtsuo, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sugabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. |  |                             |
| TITLE      | Direct Submission   |  |                             |
| JOURNAL    | Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of   |  |                             |

|  |          |  |      |  |  |
|--|----------|--|------|--|--|
| <b>COMMENT</b>   |          |  |      |  |  |
| Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216) |          |  |      |  |  |
| cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  |          |  |      |  |  |
| Please visit our web site for further details.<br>URL: http://genome.gsc.riken.jp/<br>URL: http://fantom.gsc.riken.jp/.<br>Location/Qualifiers   |          |  |      |  |  |
| <b>FEATURES</b>  |          |  |      |  |  |
| source   |          |  |      |  |  |
| 1..2841  |          |  |      |  |  |
| /organism="Mus musculus"   |          |  |      |  |  |
| /mol_type="mRNA"   |          |  |      |  |  |
| /strain="C57BL/6J"   |          |  |      |  |  |
| /db_xref="PANTOM_DB:B230378117"  |          |  |      |  |  |
| /db_xref="taxon:10090"   |          |  |      |  |  |
| /clone="B230378117"  |          |  |      |  |  |
| /sex="male"  |          |  |      |  |  |
| /tissue_type="corpora quadrigenima"  |          |  |      |  |  |
| /clone_lib="RIKEN full-length enriched mouse cDNA library"   |          |  |      |  |  |
| /dev_stage="adult"   |          |  |      |  |  |
| 1..2841  |          |  |      |  |  |
| /note="hypothetical Sushi domain / SCR repeat / CCP module/ CUB domain containing protein (InterPro IPR000436, evidence: InterPro)"  |          |  |      |  |  |
| <b>ORIGIN</b>  |          |  |      |  |  |
| <b>Alignment Scores:</b>   |          |  |      |  |  |
| Pred. No.:   | 8,89e-90 | Length:  | 2841 |  |  |
| Score:   | 1229.50  | Matches:   | 243  |  |  |
| Percent Similarity:  | 77.29%   | Conservative:  | 19   |  |  |
| Best Local Similarity:   | 71.68%   | Mismatch:  | 44   |  |  |
| Query Match:   | 7.24%    | Indels:  | 34   |  |  |
| Dbl:   | 3        | Gaps:  | 4    |  |  |
| US-10-016-248-2 (1-3104) x AK046385 (1-2841)   |          |  |      |  |  |
| OY   | 1        | MecAlaGlYAAPrOPrOAlAleuleuLeuPrOCySeRLeuLlSeSaSPCyGs           | 20   |  |  |
| Db   | 949      | CTCTGTGGTTTACCGAGCCAGCCCTCCAGCCCTGCATCAGCAGCAAGAATTGGCTGCC     | 1008 |  |  |
| OY   | 21       | AlaseRangInaGHlaserValGIyAlly-----ProSerGIueu                  | 35   |  |  |
| Db   | 1009     | GGCTTCACCTTCATCAGATGATGAACCCAGCGACGAGCTTCAGCCGCCAGTAACA-       | 1067 |  |  |
| OY   | 36       | VallYslvSGlnllEGluLeuLYseSRgSlYallyVslumePProSeZLyAsApSn       | 55   |  |  |
| Db   | 1068     | GTCAAAGAACGATTGACTGAAGTCTTGAGAGCGTGAAGCTGATGCCAGCAAGACAAAT     | 1127 |  |  |
| OY   | 56       | SerGlnYstrHserValLeuthGlnValGIyAlSeRGlngIylHIsamMetCyasPro     | 75   |  |  |
| Db   | 1128     | AACCAAGAACATCTGTGTAACTCAGGTGTGTGTCCCAAGGACATAATATGTGTCCA       | 1187 |  |  |
| OY   | 76       | ASprOGlYLlEPPrOGluarGGLylyVaRgluenuLYserAPPheaaRgluengLYserSer | 95   |  |  |
| Db   | 1188     | GACCCTGGCAATCATGAAAAAGGAAAAAGACTTAGGCTTCAGCTTCAGGTTAGATTCACGC  | 1247 |  |  |
| OY   | 96       | ValGlnPhetThyrBaSenGLUGLYTYrAsPlenGnglYserLYbaRglEtThCYemEt    | 115  |  |  |
| Db   | 1248     | GTCAAGTTCACTTCACAGAGGCGCTTAGCTGCAAGGCTCCAAAGAAATCACTGCATG      | 1307 |  |  |
| OY   | 116      | LysValSerASPmetPhealAlaITrPSerAsphIsaRgProvalCYsarGalAargMet   | 135  |  |  |
| Db   | 1308     | AAAGTGAGCGACATGTTTGCAGCATGTGAGTGAACACAGACGCGGTCTGCCAGCCGCAATG  | 1367 |  |  |
| OY   | 136      | CysaspAlahlsLemaRgGLYPrOSerGilylElleThrSerProaenPheProIlEgln   | 155  |  |  |
| Db   | 1368     | TGTATATGCCACCTTCAGAGTCCCTGGGATATCATCAGTCCCCCACTTCCCATTACG      | 1427 |  |  |

/lab host="DH10F (71 phase resistant)"  
 /clone\_lib="N1H\_BMAP\_FY0"  
 /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is AGGAGACAGAC. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemlin Chhn, Ph.D.,  
 program coordinator."



| FEATURES   | SOURCE  |
|------------|---|
| Db         | 2126 ATCTCAACGGATATCATCCAGTACAGAGGGCTTGGCTTATCTCAGACACCAACCTCTACACC 211   |
| Qy         | 2816 CybVal1ProGly1YrYrMeMeMetGlusteriariyalservalserValLeuSerCybThrLysasp 28-  |
| Db         | 2186 TCCAGGCCAGCGCTTGAAGTATAGTGGGCAATGCTACCAACCCCTC-----TGTGGGGAATAAT 222   |
| Qy         | 2836 ArgThr1TPaengly1YrYrLysProVal1CybValalaleuMetCybLysProProProLeu 285  |
| Db         | 2240 GGCAGGTGGCTCGGAGGAAAAACAATGNCMAACCATTTGAATCCAGAGCCCAAGAG 225   |
| Qy         | 2856 IleProAenglyLysValalValGlySerAspPheMet1rProLysSerVal1ThrTyrAla 287   |
| Db         | 2300 ATTTTAATGGCCCAATCTCTCTCCGTAGCGTTCTAGATAGACAAACCATCATCACTACTTT 235  |
| Qy         | 2876 CysLeuGluGly1YrYrGlnLeuSerLeuProAlaValPheThrCysGluGly1YrYrSer 289  |
| Db         | 2350 TGTATACCGGGGCTTCCGGCTCGAAGGTCCAAAATCCCTGACCTGTTAGAGACAGGTAC 241  |
| Qy         | 2896 TrpThrGlyGluLeuProGln1CysPheProVal1PheCysGlyLysApproGlyVal1ProSer 291  |
| Db         | 2420 TGGGATATGATATCCCTCTCTGTATGCCATCCATCAGTACAGTACCCACAGCCCAATTGAA 247  |
| Qy         | 2916 ArgGly1YrYrArgLysPargGlyPheSer1YrYrGlySerValserPheSerCysHis 293  |
| Db         | 2480 AATGTGTTCTGTAAAGGTGGCATATACAGATACGGTGCATATATCATCTACTCTCTTC 253   |
| Qy         | 2936 ProProLeuValLeuValGlySerProAlaGArgPheCysGlnSerAspGly1YrYrTrpSer 295  |
| Db         | 2540 CTGGGTTTCAGGTGCTTGTGTATGCGCATGCAAGACCTGTGAAGAGTCGGGA---TGGTCA 259  |
| Qy         | 2956 Gly1YrGlnProSerCysIleAspProThrLeuThr1YrYrAlaAspProGly1YrYrPro 297  |
| Db         | 2597 AGCTTCAGGCCCAACCTGTGTACCC-----ATGAGCTGGCGGTCTCCCT 263  |
| Qy         | 2976 -----GlnPheGly 2978  |
| Db         | 2639 CCTCAGATGACTTGTGT 2656   |
| RESULT 9   |   |
| LOCUS      | BX671124 673 bp mRNA linear EST 07-MAY-2004   |
| DEFINITION | BX671124 Sus scrofa 1library (scap) Sus scrofa cDNA clone   |
| ACCESSION  | scac00361.d.22 Sprim, mRNA sequence.  |
| VERSION    | BX671124  |
| KEYWORDS   | BX671124.1 GI:37982337  |
| SOURCE     | EST.  |
| ORGANISM   | Sus scrofa (pig)  |
| REFERENCE  | Sus scrofa  |
| AUTHORS    | Bukacynska, Metacoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus. 1 (bases 1 to 673)   |
| TITLE      | Bonner, A., Tosser-Klopp, G., Beme, F., Cabau, C., Villegier, S., Soares, M., Bonaldi, F. and Harey, F.   |
| JOURNAL    | A Pig Normalised Multi-Tissue cDNA Library  |
| COMMENT    | Unpublished (2003)<br>Contact: Tosser-Klopp G<br>Genetique Animale<br>Institut National de la Recherche Agronomique<br>Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan<br>Cedex, FRANCE<br>Tel: 33 (0) 5.61.28.51.14<br>Fax: 33 (0) 5.61.28.53.08<br>Email: tosser@tolouse.inra.fr<br>Sequence cleaned of vector, adaptor and repetitions. Contact us<br>at signenasupport@jouy.inra.fr to obtain the chromatogram of this<br>plate: 0036 row: d column: 22.<br>Location/Qualifiers<br>1..673<br>/organism="Sus scrofa"<br>/mol_type="mRNA"<br>/db_xref="taxon:9623" |





Oy 876 AsnTyrAsnAsnAsnH1agluCvalleTyrSer11leglnThrglnProgluYlyGly11le 895  
 Db 60 AACTACAAATTAACAACAGAGATGATCTACTCCATCCAGACCGCGGGAGAGGAGATT 119  
 Oy 896 GlnleuYsa1aargalapheserSerPhegluYsaPvalleuYsaYlyTyrAspGly 915  
 Db 120 CAGCTTAAGCAAGGATTTGAGCTGCTTAAGAGACCTCCCAAGGCTTAAGAGCGC 179  
 Oy 916 AsnAsnAsnSer1aaargleuYsa1PheSerH1sergluYsa1MetGlyVal1Thr 935  
 Db 180 ACTAACAACTCTGCCCTTGTGCTGGGGCTCTTCAGCGCTCTGAGATGTTGGGGTACT 239  
 Oy 936 LeuAsnSerThrserSerSerSerleuTripleAsnPhe11erThraP1agluAsnThSer 955  
 Db 240 TTGAACAGACAGCTTAGAGCTTG1GGCTTGACTTCAACAGATGCAAGAAACACACAGC 299  
 Oy 956 LysG1YpHeh1uYsa1PheSerSerPhegluYsa1PheYsaGluAspProgluYThr 975  
 Db 300 AAAGGCTTTGAACCTTCAAGTTCTAGTTTTCAGCTCATCCAGATGAGAGCCAGGAGAC 359  
 Oy 976 ProLysPheGlyYlyYsaValH1aPspGluYsa1PhealaglySerSerValSerPhe 995  
 Db 360 CCAAGATTGGGCTCAAGGTTCAAGATGAGGTCAATTTTCAGAGGAGACTCCGATGACCTTC 419  
 Oy 996 SerCyAsnProgluYlyTyrSerleuYsaGlySerGluYsa1PheYsaGluYsa1Glu 1015  
 Db 420 AGCTGTAACCTGGCTCAAGTCTTGGGGGAGTGAAGTCAAGATGCTGAGTGGTGA 479  
 Oy 1016 ArgArGThrTTPAaPARGProleuProThrCyVala1agluYsaGlyYlyThrValArg 1035  
 Db 480 CGCAGACGTTGAGACCCAGCCGCTCCCAACCTGTCCTGAGTGTGAGAGGAGACGTAGA 539  
 Oy 1036 GlyGluValSerGlyGlnValleuSerProgluYlyTyrProalPProYlyGlnH1aPleu 1055  
 Db 540 GGAAGGTTGTCAGAGGAGGAGTGTCTGCGCCGGGTACCCGGCTCCCTTGAACACAACTC 599  
 Oy 1056 AsnCyAlleTTPThrl1legln1aglu1aglyCySerHrl1legly 1070  
 Db 600 AACTGCACTGGACCTTGCATGAGGACGCGCGGTGACCAATTGGG 644  
 RESULT 12  
 CK12578 707 bp mRNA linear EST 01-MAR-2004  
 LOCUS SB02035A2B11.f1 normalized Keck-Tagu Library SB02 Taeniolopygia  
 DEFINITION guttata cDNA clone SB02035A2B11.f1 5, mRNA sequence.  
 ACCESSION CK12578  
 VERSION CK12578.1 GI:44822152  
 KEYWORDS EST.  
 SOURCE Taeniolopygia guttata  
 ORGANISM Taeniolopygia guttata  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;  
 Seritidae; Taeniolopygia.  
 REFERENCE 1 (bases 1 to 707)  
 AUTHORS Clayton, D.F., Arnold, A.P., Ball, G.F., Brenowitz, E., George, J.M.,  
 Mello, C.V., Wade, J., Replogle, K., Lewin, H., Band, M., Hernandez, A.,  
 and Liu, L.  
 TITLE The Songbird Neurogenomics Initiative: An Evolving Public Resource  
 JOURNAL for Study of Genes, Brain, and Behavior  
 COMMENT Unpublished (2004)  
 Contact: David F. Clayton  
 University of Illinois  
 B107 CLSL, 601 S. Goodwin, Urbana, IL 61801, USA  
 Tel: 217 244 3668  
 Fax: 217 244 1648  
 Email: dclayton@uiuc.edu  
 Base Calling/Quality Scores: PHRED from Washington University  
 Genome Center.  
 Vector Trimming: Cross\_match from Washington University Genome  
 Center PHAP suite. Low quality bases (Phred score < 20) were  
 trimmed from both ends of the sequence by an in-house script.  
 This sequence is vector free and at least 200 bp in length. Funded

by PHS grant # R01 NS045264, 'Songbird Neurogenomics Initiative.'  
 PCR Primers  
 FORWARD: TAATACGACTCACTATAGG (T)  
 BACKWARD: ATTAACCTCACTAAAG (T)  
 Insert Length: 707 Std Error: 0.00  
 Plate: SB02035A2 row: B column: 11  
 Seq primer: TAATACGACTCACTATAGG (T)  
 High quality sequence stop: 707.  
 Location/Qualifiers  
 1..707  
 /organism="Taeniolopygia guttata"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:59729"  
 /clone="SB02035A2B11.f1"  
 /tissue\_type="brain"  
 /dev\_stage="late embryo, post-hatch days 1, 10, 20, 45,  
 and adult (pooled)"  
 /lab\_host="DH10B"  
 /clone\_1lb="normalized Keck-Tagu Library SB02"  
 /notes="Organ: brain; Vector: PBS II SK(+); Site: 1;  
 EcorI(5' side of insert); Site 2: NotI (3' side of  
 insert); The library was constructed and normalized as  
 described by Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 (1996), Genome Research 6(9): 791-806. An identifying tag  
 was added at the 3' during cDNA synthesis:  
 insertCAAAAAAAAAAAAAAAAAATGCA."  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,01e-76 Length: 707  
 Score: 1053.00 Matches: 190  
 Percent Similarity: 93.43% Conservative: 9  
 Best Local Similarity: 89.20% Mismatches: 14  
 Query Match: 6.20% Indels: 0  
 DB: 7 Gaps: 0  
 US-10-016-248-2 (1-3104) x CK12578 (1-707)  
 Oy 1514 A1a1eVala1argPhegluYsa1AsnSerGlyTyrAla1euglnGlySerProglu11leGlu 1533  
 Db 3 GGNNNNNNGCGCTTCAGTGCAGACCGCGCTTACCGCTGGAGAGGCTCCCTGCATCAG 62  
 Oy 1534 CyAsnProValProgluYsa1a1eVala1ag11trPasnValSerAlaProThrCyVala1Val 1553  
 Db 63 TGCTTGCCACATGCCCGGGGCTTGGCAAGTGAACCTCCGCTCCACCTGGTGTG 122  
 Oy 1554 ProCyGlyYsa1e1uThrgluYsa1argGlyThrl1e1uSerProgluYsa1PheProglu 1573  
 Db 123 CCTGCGGGGGGAACCTGACGAGAGAGAAAGAACATCTCTGCCAGGCTTCCGTGAG 182  
 Oy 1574 ProTyrleuAsnSerleuAsnCyVala1TTPYsa11eVala1ProgluGlyAlaGly11le 1593  
 Db 183 CCTTACCTGAACAGCTTCAACCTGCTGTGGAAGTCAAGTGCCTCCGAGGAGCGGATC 242  
 Oy 1594 Gln11leglnVala1SerPheValThrgluGluAsnThraPspSerleuGluValPheAsp 1613  
 Db 243 CAGATCCAGGTGATGATGATTTGTTCAGAGACAGACTGGGATTTCCCTGGAAGTGTGAT 302  
 Oy 1614 GlyAlaAsnAsnThrVal1ThrMetleuGlySerPheSerGlyThrThrValProal1eU 1633  
 Db 303 GAGAGAGAACAAACAGACCAATGCTGGGAGGCTTCTGGAACCAAGTGCCTGCGCTG 362  
 Oy 1634 LeuAsnSerThrserAsnGlnleuYlyYsa1PheYlyserAsp11eSerValSerAla 1653  
 Db 363 CTCAACAGCACTTCAACCAAGCTTATCTATCTCAATTCGACATCAGCTGCTCGCT 422  
 Oy 1654 AlaGlyPheH1sleuGluYlyYsa1ThrValGlyleuSerSerCyPspProAlaVal 1673  
 Db 423 GCGGCTTTTCACTGGAATTCAGAAACCGTGGGCTTCTCCAGCTCCGAGACCGGCGCTT 482  
 Oy 1674 ProSerAsnGlyValYlyThrglyYsa1argTyrleuVala1AsnAspVala1SerPheGln 1693  
 Db 483 CTTGGAATGGGCTGGAAGATCGGGAGCGGCTACTGTTGAACGAGCGTGTCTTCCAG 542

|   |   |   |                             |
|---|---|---|-----------------------------|
| Qy  | 1694  | CyAGlucProglYThyrAlaLeuGlnGlyYHAlaHisIleSerCYweIProglYThyrVal | 171311                      |
| Db  | 543   | TCGAGAGCCCGGCTACGCGCTCGACGGGCGACCTCCCATCTCTCGATGCGCGGCGACGGTG | 602                         |
| Qy  | 1714  | ArgArgTTPAsnTYrProProProlaCYsIleAlaGln                        | 1726                        |
| Db  | 603   | CGCCGCTGGAAATTAACCTCCACCCCTCTCGATCGGTAGG                      | 641                         |
| RESULT 13                                   |   |   |                             |
| LOCUS                                       | AL638364  | 661 bp  | mRNA linear EST 07-NOV-2003 |
| DEFINITION                                  | AL638364 XCC-neurula Xenopus tropicalis cDNA clone TNeu018n21 5', mRNA sequence.  |   |                             |
| ACCESSION                                   | AL638364  |   |                             |
| VERSION                                     | AL638364.2  | GI:38217218   |                             |
| KEYWORDS                                    | EST.  |   |                             |
| SOURCE                                      | Xenopus tropicalis (western clawed frog)  |   |                             |
| ORGANISM                                    | Xenopus tropicalis  |   |                             |
| REFERENCE                                   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.   |   |                             |
| AUTHORS                                     | 1 Croising,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.   |   |                             |
| TITLE                                       | Sanger Xenopus tropicalis EST project 2001 (11_2003)  |   |                             |
| JOURNAL                                     | Unpublished (2003)  |   |                             |
| COMMENT                                     | On Nov 7, 2001 this sequence version replaced gi:16790343.<br>Contact: Hockle E<br>Sanger Institute<br>Hinxton, Cambridgeshire, CB10 1SA, UK<br>Email: trop@sanger.ac.uk<br>This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.<br>cDNA was oligo dt primed from 5ug of poly A+ RNA from neurula.<br>ECORI-NotI cut cDNA was then ligated into pCS107 with EcorI at the 5' end and NotI at the 3' end.<br>Vector: pCS107; Site_1: EcorI; Site_2: NotI<br>Host: Escherichia coli DH10B<br>Sanger Xenopus tropicalis EST project 2001<br>TROPICALIS SEQUENCE ID: TNeu018n21.plksp6<br>Sequencing primer: SP6<br>Location/Qualifiers<br>1..661<br>/organism="Xenopus tropicalis"<br>/mol_type="mRNA"<br>/db_xref="taxon:8364"<br>/clone="TNeu018n21"<br>/dev_stage="neurula"<br>/lab_host="Escherichia coli DH10B"<br>/clone_lib="XCC-neurula"<br>/note="Vector: pCS107; Site_1: EcorI; Site_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from neurula.<br>ECORI-NotI cut cDNA was then ligated into pCS107 with EcorI at the 5' end and NotI at the 3' end." |   |                             |
| FEATURES                                    |   |   |                             |
| source                                      |   |   |                             |
| ORIGIN                                      |   |   |                             |
| Alignment Scores:                           |   |   |                             |
| Prod. No.:                                  | 8.03e-73  | Length:   | 661                         |
| Score:                                      | 1011.00   | Matches:  | 174                         |
| Percent Similarity:                         | 88.58%  | Conservative:   | 20                          |
| Best Local Similarity:                      | 79.45%  | Mismatches:   | 25                          |
| Query Match:                                | 5.95%   | Indels:   | 0                           |
| DB:   | 1   | Gaps:   | 0                           |
| US-10-016-248-2 (1-3104) x AL638364 (1-661) |   |   |                             |
| Qy  | 2800  | LeuArgLeuGlyValAspPheArgTYrAsnTYrThValThTYrGlnCYsValProGly    | 2819                        |
| Db  | 2   | CTACGGCTGCACCATGATTTCCAGTCAACACACAGTGCCTTCACAGTTTGCTGGA       | 61                          |
| Qy  | 2820  | TYMeMeGluSerHisArgValSerValIleuSerCYsThLYsAspArgThTTPAsn      | 2839                        |
| Db  | 62  | TATATATGGAATCAACAAAGGACCCATGTTACTGTGACCAAAAGACGTACCTGGAA      | 121                         |

[illegible]



/clone.lib="Sus Scrofa library (scac)"  
 /note="Vector: pT73D-pac vector; tissues: adipose tissue,  
 brain, kidney, liver, muscle, ovary, testis, heart,  
 hypothalamus, pancreas, skin, spleen, thymus, placenta,  
 pituitary gland, seminal vesicle, small intestine,  
 uterus, adrenals, bulbo urethral gland, cerebral trunk,  
 epididymis, female gonad, gall-bladder, hippocampus,  
 large intestine, male gonad, melanocytes, stomach, udder"

## ORIGIN

## Alignment Scores:

|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 2,396-66 | Length:       | 671 |
| Score:                 | 933.00   | Matches:      | 165 |
| Percent Similarity:    | 98.83%   | Conservative: | 4   |
| Best Local Similarity: | 96.49%   | Mismatches:   | 2   |
| Query Match:           | 5        | Indels:       | 0   |
| DB:                    | 5        | Gaps:         | 0   |

US-10-016-248-2 (1-3104) x BX676029 (1-671)

```

QY      1727  CysGlyGlyThrValGluGluMetGluGlyValIleLeuSerProGlyPheProGlyAsn 1746
DB      3      TGTGGGGAGAGAGTGGAAAGACATGAGGGGGGTGATCTGAGCCCGGCTTCCAGGCAAC 62
QY      1747  TyrProSerAsnMetAspCysSerTrpIleAlaLeuProValGlyPheGlyAlaHis 1766
DB      63      TACCCACAGCAATGAGACTGTTCCTGGAAATAGCCCTGCCGTGGGCTTTGGAGCTCAC 122
QY      1767  IleGlnPheLeuAsnPheSerThrGluProAsnHisAspTrpIleGluIleArgAsnGly 1786
DB      123  ATCCAGTTCCTGAACCTTCCACGAGGCCAACCATGATCATCGAAATCCGGAATGGC 182
QY      1787  ProTyrGluThrSerArgMetMetGlyArgPheSerGlySerGluLeuProSerSerLeu 1806
DB      183  CCTACGAGACCAAGCCGATGATGAGGAGATTCAAGCGAGAGAGCTGCCAGGCCCTG 242
QY      1807  LeuSerThrSerHisGluThrThrValIleTyrPheHisSerAspHisSerGlnAsnArgPro 1826
DB      243  CTCTCCAGTCCACGAGACCAACCGTGTACTTCCACAGCACCACTCCAGAACCGGCGC 302
QY      1827  GlyPheLeuLeuGluTyrGlnAlaTyrGluLeuGlnGluCysProAspProGluProPhe 1846
DB      303  GGATTCAAGCTGGAGTACAGGCGCTATGAATTCAGAGTGGCCAGACCAAGACCTTT 362
QY      1847  AlaAsnGlyIleValArgGlyAlaGlyTyrAsnValGlyGlnSerValThrPheGluCys 1866
DB      363  GCCAATGACATCGTAAAGGGGTGCTGCTACACGTTGGCAGTCTGAACCTTCGAGTGC 422
QY      1867  LeuProGlyTyrGlnLeuThrGlyHisProValLeuThrCysGlnHisGlyThrAsnArg 1886
DB      423  CTCCGGGGGTATCAACTGATGAGGGCCACCTGTTCCTCATGATCAGACGCGACCAACCG 482
QY      1887  AsnTrpAspHisProLeuProGlyCysGluVal 1897
DB      483  AACTGGAGCACCCCTGCGCCAGGTGTGAAGTG 515

```

Search completed: October 18, 2004, 15:00:36  
 Job time : 18322 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus model

Run on: October 18, 2004, 09:55:06 ; Search time 2743 Seconds  
(without alignments)  
5940.285 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985  
Sequence: 1 MAGAPPPALLPCLSLSDCC.....RSGPVGDPSTLPQSHRSPKP 3104

Scoring table:

BLISSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h  
-Q/cg2\_1/USPTO.spool/US10016248/runat.18102004.105458.11519/app.query.fasta\_1.3271  
-DB=N\_Geneseq.23Sep04 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blissum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10016248 @CGN\_1\_1\_1736 @runat.18102004.105458.11519 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq.23Sep04:\*  
1: geneeqn1980s:\*  
2: geneeqn1990s:\*  
3: geneeqn2000s:\*  
4: geneeqn2001as:\*  
5: geneeqn2001bs:\*  
6: geneeqn2002as:\*  
7: geneeqn2002bs:\*  
8: geneeqn2003as:\*  
9: geneeqn2003bs:\*  
10: geneeqn2003cs:\*  
11: geneeqn2003ds:\*  
12: geneeqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| No. | Score | Query Match | Length | ID | Description        |
|-----|-------|-------------|--------|----|--------------------|
| 1   | 16985 | 100.0       | 10136  | 6  | AB664375 Human cub |
| 2   | 16985 | 100.0       | 10136  | 12 | ADH71165 Human gen |
| 3   | 16985 | 100.0       | 10136  | 12 | ADH71167 Human gen |
| 4   | 16970 | 99.9        | 10136  | 12 | ADH71167 Human gen |
| 5   | 16512 | 97.2        | 9951   | 12 | ADH71141 Human gen |
| 6   | 16319 | 96.1        | 10655  | 12 | ADH71135 Human gen |

|    |         |      |       |    |          |                    |
|----|---------|------|-------|----|----------|--------------------|
| 7  | 15917.5 | 93.7 | 10466 | 12 | ADH71143 | Adh71143 Human gen |
| 8  | 15519   | 91.4 | 12900 | 12 | ADH71145 | Adh71145 Human gen |
| 9  | 14142.5 | 83.3 | 8010  | 6  | AB664375 | AB664375 Human cub |
| 10 | 14142.5 | 83.3 | 8010  | 12 | ADH71139 | Adh71139 Human gen |
| 11 | 11324   | 66.7 | 10989 | 12 | ADH72215 | Adh72215 Human gen |
| 12 | 10780   | 63.5 | 10433 | 6  | AAD33320 | Aad33320 Human C3b |
| 13 | 10737   | 63.2 | 10673 | 6  | AAD33318 | Aad33318 Human C3b |
| 14 | 10720   | 63.1 | 12525 | 6  | AAD33319 | Aad33319 Rat C3b/C |
| 15 | 9077.5  | 53.4 | 6004  | 6  | AA149944 | AA149944 Human mol |
| 16 | 7643.5  | 45.0 | 7323  | 6  | AA188805 | AA188805 DNA encod |
| 17 | 7643.5  | 45.0 | 8034  | 6  | AA188806 | AA188806 DNA encod |
| 18 | 7610.5  | 44.8 | 6409  | 6  | AA188802 | AA188802 DNA encod |
| 19 | 7159    | 42.1 | 5598  | 6  | AA188803 | AA188803 DNA encod |
| 20 | 6991    | 41.2 | 6145  | 6  | AA188802 | AA188802 DNA encod |
| 21 | 6978    | 41.1 | 5667  | 6  | AA188804 | AA188804 DNA encod |
| 22 | 5172.5  | 30.5 | 3896  | 4  | AA188804 | AA188804 NOV16 cod |
| 23 | 5172.5  | 30.5 | 3896  | 12 | ADH71163 | Adh71163 Human gen |
| 24 | 5129    | 30.2 | 3904  | 12 | ADH71161 | Adh71161 Human gen |
| 25 | 5129    | 30.2 | 3905  | 4  | AA188802 | AA188802 NOV15 cod |
| 26 | 4864    | 28.6 | 2897  | 10 | AD30517  | Ad30517 Human nov  |
| 27 | 4692    | 27.6 | 2607  | 12 | ADH71153 | Adh71153 Human gen |
| 28 | 4268    | 25.1 | 2328  | 12 | ADH71149 | Adh71149 Human gen |
| 29 | 4257    | 25.1 | 2329  | 12 | ADH71147 | Adh71147 Human gen |
| 30 | 4148.5  | 24.4 | 4506  | 6  | AB58378  | AB58378 Protein m  |
| 31 | 3942    | 23.2 | 2487  | 4  | AA500146 | AA500146 Human CDN |
| 32 | 3856.5  | 22.7 | 2139  | 12 | ADH71151 | Adh71151 Human gen |
| 33 | 3239    | 19.1 | 1792  | 12 | ADH71157 | Adh71157 Human gen |
| 34 | 2894.5  | 16.8 | 3019  | 12 | ADH72219 | Adh72219 Human gen |
| 35 | 2694    | 15.9 | 2609  | 6  | AAD37600 | Aad37600 Human int |
| 36 | 2685    | 15.8 | 2026  | 10 | AD30413  | Ad30413 Human nov  |
| 37 | 2639    | 15.5 | 2387  | 6  | ABN99361 | ABN99361 Human sec |
| 38 | 2025.5  | 11.9 | 1749  | 10 | AD30885  | Ad30885 Human nov  |
| 39 | 1940.5  | 11.4 | 1615  | 6  | ABA00063 | ABA00063 CADHP-10  |
| 40 | 1852    | 10.9 | 11230 | 6  | AAD32026 | Aad32026 Mouse C3b |
| 41 | 1797.5  | 10.6 | 10878 | 6  | AAD32025 | Aad32025 Human C3b |
| 42 | 1796.5  | 10.6 | 11158 | 6  | ABT08490 | ABT08490 Human nov |
| 43 | 1796.5  | 10.6 | 11158 | 12 | ADH71299 | Adh71299 Human gen |
| 44 | 1796.5  | 10.6 | 11158 | 12 | ADH71301 | Adh71301 Human gen |
| 45 | 1796.5  | 10.6 | 11158 | 12 | ADH71303 | Adh71303 Human gen |

#### ALIGNMENTS

|          |  |
|----------|--|
| RESULT 1 |  |
| ID       | AB664375 standard; DNA, 10136 BP.  |
| XX       | AB664375;  |
| AC       |  |
| XX       |  |
| DT       | 15-NOV-2002 (first entry)  |
| XX       |  |
| DE       | Human cub and sushi domain containing gene #1.                           |
| XX       |  |
| KW       | Human; NOV5; neurodegenerative disease; Alzheimer's disease; anxiety;    |
| KW       | Parkinson's disease; Huntington's disease; neurological disorder;        |
| KW       | schizophrenia; manic depression; mental retardation; angina pectoris;    |
| KW       | cardiovascular disease; acute heart failure; myocardial infarction;      |
| KW       | muscular disease; muscular disorder; retinal disease; photoreception;    |
| KW       | deafness; keratinization disorder; cancer; ovarian cancer; melanoma;     |
| KW       | immunological disorder; inflammatory disease; immune disease; diabetes;  |
| KW       | bacterial infection; fungal infection; protozoal infection; obesity;     |
| KW       | viral infection; reproductive system disorder; metabolic disturbance;    |
| KW       | anorexia; wasting disorder; chronic disease; infectious disease;         |
| KW       | dyslipidemia; cub; sushi; myelin; von willebrand factor; kiellin;        |
| KW       | semaphorin; serine/threonine protein kinase; TGF-beta binding;           |
| KW       | mas proto-oncogene; ribonuclease pancreatic precursor; aminotransferase; |
| KW       | toll-like 2; cysteine sulfinic acid decarboxylase; gene; ds;             |
| KW       | single nucleotide polymorphism; SNP.                                     |
| XX       |  |
| XX       | Homo sapiens.  |
| OS       |  |
| XX       |  |
| XX       |  |
| FT       | Key Location/Qualifiers  |
|          | variation replace(3082..A)   |

|                   |   |   |
|-------------------|---|---|
| FT                | /tag= a   | /note= "Single nucleotide polymorphism (SNP)" |
| FT                | replace(3120,A)   |   |
| FT                | /tag= b   |   |
| FT                | /note= "Single nucleotide polymorphism (SNP)"                             |   |
| FT                | replace(3251,C)   |   |
| FT                | /tag= c   |   |
| FT                | /note= "Single nucleotide polymorphism (SNP)"                             |   |
| FT                | replace(4085,G)   |   |
| FT                | /tag= d   |   |
| FT                | /note= "Single nucleotide polymorphism (SNP)"                             |   |
| PN                | MO200264791-A2.   |   |
| XX                |   |   |
| XX                | 22-AUG-2002.  |   |
| XX                |   |   |
| XX                | 10-DEC-2001; 2001MCO-US048369.  |   |
| XX                |   |   |
| PR                | 08-DEC-2000; 2000US-02554329P.  |   |
| PR                | 14-DEC-2000; 2000US-0255648P.   |   |
| PR                | 15-MAY-2001; 2001US-02911037P.  |   |
| PR                | 08-JUN-2001; 2001US-02971173P.  |   |
| PR                | 08-JUN-2001; 2001US-0309258P.   |   |
| PR                | 29-AUG-2001; 2001US-0315639P.   |   |
| PR                | 01-OCT-2001; 2001US-0326393P.   |   |
| XX                |   |   |
| PA                | (CUBA-) CUBAGEN CORP.   |   |
| PI                | Alsbrook JP, Anderson DW, Burgess CE, Boldog FL, Casman SJ;               |   |
| PI                | Colman SD, Edinger SR, Ellemann K, Gerlach V, Gorman L, Grosse WM,        |   |
| PI                | Guo X, Herrmann JL, Kehda R, Lepley DM, Li H, MacDougall JR,              |   |
| PI                | Millet I, Pena CE, Peyman JA, Rastelli L, Rieger DK, Shinkets RA;         |   |
| PI                | Smithson G, Spytek KA, Stone DJ, Tchervet VT, Vernet CM, Voss EZ;         |   |
| PI                | Zernhusen BD, Zhong H, Zhong M;   |   |
| XX                |   |   |
| XX                | WPI; 2002-643486/69.  |   |
| XX                | P-PSDB; ABG79168.   |   |
| PT                | New NOXV polypeptides and polynucleotides useful for treating or          |   |
| PT                | preventing e.g. neurodegenerative diseases, neurological disorders,       |   |
| PT                | cardiovascular diseases, muscular diseases and disorders, or              |   |
| PT                | immunological diseases.   |   |
| XX                |   |   |
| XX                | Claim 9; Page 10-12; 29pp; English.                                       |   |
| XX                |   |   |
| CC                | The present invention relates to new NOXV polypeptides. The polypeptides, |   |
| CC                | polynucleotides and antibodies are useful in the manufacture of a         |   |
| CC                | medicament for treating or preventing neurodegenerative diseases (e.g.    |   |
| CC                | Alzheimer's disease, Parkinson's disease, or Huntington's disease),       |   |
| CC                | neurological disorders (e.g. anxiety, schizophrenia, manic depression or  |   |
| CC                | mental retardation), cardiovascular disease (e.g. acute heart failure,    |   |
| CC                | angina pectoris or myocardial infarction), muscular diseases and          |   |
| CC                | disorders, retinal diseases (including those involving photoreception,    |   |
| CC                | deafness and keratinisation disorders), cancer (e.g. ovarian cancer or    |   |
| CC                | melanoma), immunological disorders, inflammatory and immune diseases,     |   |
| CC                | bacterial, fungal, protozoal and viral infections, and reproductive       |   |
| CC                | system disorders. The proteins of the invention may be used to screen     |   |
| CC                | drugs or compounds that modulate the NOXV protein activity or expression, |   |
| CC                | as well as to treat disorders characterised by insufficient or excessive  |   |
| CC                | production of NOXV protein or protein forms that have decreased or        |   |
| CC                | aberrant activity compared to NOXV wild type protein, such as diabetes,   |   |
| CC                | obesity, metabolic disturbances associated with obesity, anorexia and     |   |
| CC                | wasting disorders associated with chronic diseases and various cancers,   |   |
| CC                | infectious diseases and various dyslipidaemias. The nucleic acid          |   |
| CC                | sequences of the invention may be used in chromosome mapping, identifying |   |
| CC                | an individual from minute biological samples (tissue typing), and in      |   |
| CC                | forensic identification of a biological sample. The present nucleic acid  |   |
| CC                | sequence encodes a NOXV protein of the invention                          |   |
| XX                |   |   |
| XX                | Sequence 10136 BP; 2147 A; 3060 C; 2744 G; 2185 T; 0 U; 0 Other;          |   |
| Alignment Scores: | 0   | Length: 10136                                 |
| Pred. No.:        |   |   |

| Score:  | 16985.00 | Matches:  | 3104 |
|---|----------|---|------|
| Percent Similarity:                           | 100.00%  | Conservative:   | 0    |
| Best Local Similarity:                        | 100.00%  | Mismatch:   | 0    |
| Query Match:                                  | 100.00%  | Indels:   | 0    |
| DB:   | 6        | Gaps:   | 0    |
| US-10-016-248-2 (1-3104) x ABSE4375 (1-10136) |          |   |      |
| QY  | 1        | MeCAlaGlyAlaProPProAlaLeuLeuProCysSerLeuIleSerAspCys          | 20   |
| Db  | 1        | ATGGCGGCGCCCTCCCGCCCTGCTGCTGCTGCAATTGATCTCAGACTGCTGT          | 60   |
| QY  | 21       | AlaSerAsnGlnArgHisSerValGlyValGlyProSerGluLeuValLysGlnIle     | 40   |
| Db  | 61       | GCTACGACATACGGACACTCTCGTGGCGTATGAGACCTCCGAGCTAGTCAAGAACCAATT  | 120  |
| QY  | 41       | GluLeuIysSerArgGlyValLysLeuMetProSerLysAspAsnSerGlnLysThrSer  | 60   |
| Db  | 121      | GAGTTAAGCTCGAGGTGTGAAGCTGATGCCAGCAAGACCAAGCCAGAAAGACTCT       | 180  |
| QY  | 61       | ValLeuThrGlnValGlyValSerGlnGlyHisAsnMetCysProAspProGlyLlePro  | 80   |
| Db  | 181      | GTTTAACTCAGGTGTGTGTGCTCCAGAGACATAAATATGTCTCAGACCTGCGCATACC    | 240  |
| QY  | 81       | GluArgGlyLysArgLeuGlySerAspPheArgLeuGlySerSerValGlnPheThrCys  | 100  |
| Db  | 241      | GAAGGGGCAAAAGACTGAGCTCGATTTCAGTTAGATTCAGCGTCCAGCTTCACCTGC     | 300  |
| QY  | 101      | AsnGluGlyThrAspLeuGlnGlySerLysArgLleThrCysMetLysValSerAspMet  | 120  |
| Db  | 301      | AACGAGGGCTATACACTGGAAGGGTCCAAAGCGGATCACTGTATGAAGAGCGACATG     | 360  |
| QY  | 121      | PheAlaAlaTrpSerAspHisArgProValCysArgAlaArgMetCysAspAlaHisLeu  | 140  |
| Db  | 361      | TTTGGGGCTGGAGGACACACAGGCCAGCTCCGAGCCCGCATGTGTATGCCACTT        | 420  |
| QY  | 141      | ArgGlyProSerGlyLleIleThrSerProAsnPheProIleGlnLysAspAsnAla     | 160  |
| Db  | 421      | CGAGGCCCTCGGGCATCATCACTCCCGCAATTCCTCATTCAGTATGACATCAACATGCA   | 480  |
| QY  | 161      | HisCysValTrpGlyIleThrAlaLeuAsnProSerLysValLleLysLeuAlaPheGln  | 180  |
| Db  | 481      | CACGTGTGTGATCATCATCAGCACTCAACCCCTCCAGAGTATCAAGCTCGCTTTGAG     | 540  |
| QY  | 181      | GluPheAspLeuGluArgGlyLysThrAspThrLeuThrValGlyAspGlyGlnAspGly  | 200  |
| Db  | 541      | GAGTTGATTGAGAGGGGCTATACACCTCGAGCGGTGTGATGTGTGTCAGATGGG        | 600  |
| QY  | 201      | AspGlnLysThrValLeuLysMetSerGlnAsnAlaCysSerAspSerProHisThrPro  | 220  |
| Db  | 601      | GACCGAAAGACAGTTCTCTCACTGCTCAAAATGGCTGTGAGTACAGCCCTCACACCCA    | 660  |
| QY  | 221      | GlySerArgLleProGluSerMetSerGlyAspLleTrpArgGlnLysTrpThrValLeu  | 240  |
| Db  | 661      | GGGCTCTCGCATCCAGAGAGCATGTCTGGGGGACATCTGGAGCCAAATGACTGTACTT    | 720  |
| QY  | 241      | GluLleCysArgAspLleSerSerSerAspAlaArgSerGlySerValArgLysSerPro  | 260  |
| Db  | 721      | GAGATCTGTGTGATTAAGCATTAAGCATTCACATCCAGGTTCAGTAGGAAAGTCTCCA    | 780  |
| QY  | 261      | LysThrSerAsnAlaValGluLeuValAlaProGlyThrGluLleGlnGlnLysSerCys  | 280  |
| Db  | 781      | AAGACTTATATGCTGTGAGACTTGTCTCTCTGGACAGAGATGAGACAGGGCAGTTGC     | 840  |
| QY  | 281      | GlyAspProGlyLleProAlaLysGlyArgGlyLysSerArgPheHisIleGlyAsp     | 300  |
| Db  | 841      | GGTAGCCCTGGCAATACCTGATATGCGCGGAGGGAGAGCGCTCCGGTTTACCAACGGTGAC | 900  |
| QY  | 301      | ThrLeuLysPheGlnCysGlnProAlaPheGluLeuValGlyGlnLysValIleThrCys  | 320  |
| Db  | 901      | ACACTCAAGTTGATGTGCGCACCCCGCTTTGATGCTGGGGACAGAAAGCATATCATGTC   | 960  |
| QY  | 321      | GlnLysAsnAsnGlnTrpSerAlaLysLysProGlyCysValPheSerCysPhePheAsn  | 340  |

```
Db 961 CAAAAGATACCAATGGTGGCTTAAGAGCCAGGCTGGCTTCTCTGCTTCTTCAAC 1020
Qy 341 PheTherSerProSerGlyValValLeuSerProAenTyProGluAerTyGlyAAsnHis 360
Db 1021 TTCCACGAGCCGCTGGGGTGTCTCTCTCCCAATCCAGAGGACTTAAGGCAACAC 1080
Qy 361 LeuHisCyValTTrPLeuHisLeuAlaArProGluSerArgHisLeuAlaPheAsn 380
Db 1081 CTCACATGTGTCTGAGCTCATCTGCGCCAGGCTTGAAGCCGACCTTCACTGAGCTTCAAC 1140
Qy 381 AspLeuAspValGluProGluPheAspPheLeuValLeuAspGlyAlaThrAlaGlu 400
Db 1141 GACATTGACGTGGAGCCCTCACTTGTATTCTCTGTCATCAAGATGGGGCCACCGCCGAG 1200
Qy 401 AlaProValLeuGlyThrPheSerGlyAAsnGluLeuProSerSerHisTherSerGly 420
Db 1201 GCGCCCTCTCCGCGACCTTCTCAGGAAACCAAGCTTCTCTCATCACAAAGCAGTGGC 1260
Qy 421 HisValAlaArgLeuGluPheGluThrAspHisSerThyGlyAspArgGlyPheAsnHis 440
Db 1261 CACGTGGCCGCTCTGAGTTCCAGATTCAGACCTGACCACTCCAGGAGAGAGGAGGCTTCAACATC 1320
Qy 441 ThrPheThrThrPheArgHisAAsnGlyCyAspProAspProGlyValProValAAsnGlyAs 460
Db 1321 ACTTTTACCACTTCCGACCAACAGAGTGGCCGATCCTGGCGTTCCAGTAAATGGCAAA 1380
Qy 461 ArgPheGlyAspSerLeuGluLeuGlySerSerHisSerPheLeuCyAspGluGlyPhe 480
Db 1381 CGGTTTGGGGACACCTTCCAGCTGGGCGAGCTCCATCTCTCTCTGTAAGAGGCTTC 1440
Qy 481 LeuGlyThrGlnGlySerGluThrHisCysValLeuGlyGluGlySerValValTTrp 500
Db 1441 CTTGGAGCTCAGGGCTCAGAGACCAATCACTGCGCTCTGAGAGAGGAGGAGGCTGTGG 1500
Qy 501 AsnSerAlaValLeuArgCyAspGluAlaProCyAspGlyGlyHisLeuTherSerProSerGly 520
Db 1501 AACAGCCGCTGTGCGGCTGAGAGCTCCCTGTGGTGTACCTGACTGCCCAAGGCGC 1560
Qy 521 ThrHisLeuSerProGlyTTrpProGlyPheThyGlyAspAlaLeuSerCyAspAlaTTrpVal 540
Db 1561 ACCATCTCTCTCTCGGGCTGGCTGGCTTCTTCAAGAGATGCTTGAAGTGGCTGGGTTG 1620
Qy 541 HisGluAlaGlnProGlyTTrpProLeuValHisLeuTherAspArgPheThyGluVal 560
Db 1621 ATTAAAGCCAGCCAGGCTACCCCAATCAAAATCACTTCCAGACATTCAAAACGAGGCTC 1680
Qy 561 AsnTyAspThrLeuGluValArgAspGlyArgThyThySerAlaProLeuHisGlyVal 580
Db 1681 AACTATAGACACCTCGGAAGTACGGAGTGGGCGGACTTACTCAGGCGCCCTGATGGGGTT 1740
Qy 581 ThyHisGlyThrGlnValProGluPheLeuHisSerTherSerAenTyLeuTyLeuLeu 600
Db 1741 TACACCGGAGCCCGAGTTCCTCCACTTCTCATCAGACACAGCACTTACTTACTCTCTC 1800
Qy 601 PheSerThrAspLeuSerHisSerAspHisLeuGlyPheGluLeuArgGlyThrHisLeuTh 620
Db 1801 TTCTCTACCGACAGAGTCACTCGACATCGGCTTCCAGCTCCGCTATGAGACTATACCA 1860
Qy 621 LeuGlnSerAspHisCysLeuAspProGlyHisProValAAsnGlyGlnArgHisGlyAsn 640
Db 1861 CTGCAAGTCAAGCACTGTCTGGATCCAGGAATCCCAAGTAAATGACAGCGTCAAGGAAAT 1920
Qy 641 AspPheThyValGlyAlaLeuValThrPheSerCyAspSerGlyTyThrTherSerAsp 660
Db 1921 GACTTCTTACGTTGGGCGGCTGGTACCTTCACTGTGACTCGGGCTACACTTAAGTGAAC 1980
Qy 661 GlnValProLeuGluGlyGluPheGluProAspPheGluThrPheArgAlaLeuProSerCyGlu 680
Db 1981 GGGAGGCTCTGAGTGTGAGCCCAACTTCCAGTGGAGCCGGGCGCTGCGCACTTGTGA 2040
Qy 681 AlaLeuCyAspGlyPheHisLeuGlnGlySerSerGlyThrHisLeuSerProGlyPhePro 700
Db 2041 GCTCTGTGTGGCTTCACTTCAAGGCTCCAGTGGAGCAATCTTGTGCCAGAGGCTTCC 2100
Qy 701 AspPheThyProAsnLeuAsnCyThrThrHisLeuHisSerHisGlyValGly 720
Db 2101 GACTTCTTACCCCAACCACTTGAAGTCACTGATTTATGAAATCTCATGAGAGGCT 2160
Qy 721 ValPhePheThrPheHisThrPheHisLeuGluSerGlyHisAspTyLeuLeuHisLeuTh 740
Db 2161 GTGTTCTTCACTTCCACACTTCCACTGGAAGAGTGGCATGACTTCTCTCATCTACT 2220
Qy 741 GluAsnGlySerPheThrGlnProLeuArgGluLeuThrGlySerAlaLeuProAlaPro 760
Db 2221 GAGAACGGAGCTTCAACCCAGCCCTGAGGAGCACTGAATCGGCTGCGCCAGCTCCC 2280
Qy 761 HisSerAlaGlyLeuTyThyGlyAsnPheThrHisGlnValArgPheHisSerAspPheSer 780
Db 2281 ATCAGCGCTGGGCTTATGCAACTTCACTGAGCCAGGCTGGCTTCACTGTGATTTCTC 2340
Qy 781 MetSerTyGluGlyPheAsnHisLeuThrPheSerGlyTyAspLeuGluProCyAspGluGlu 800
Db 2341 ATGTCATATGAGATTCACATCACTCTTCAAGTACGACTTGGAGCCCTGAGAGAG 2400
Qy 801 ProGluValProAlaTySerHisLeuArgGlyLeuGlnPheGlyValGlyAspTherLeu 820
Db 2401 CCGAGAGTCCAGGCTACACATCCGGAAGGCTTGCAGTTGGCTGGCGGACACCTTG 2460
Qy 821 ThrPheSerCyAspProGlyTTrpArgLeuGluGlyThrHisLeuArgHisLeuGly 840
Db 2461 ACCTCTCTCTCTCTCCGCGTACCTGTGAGAGGACCCCGCCAGTACAGTGTGGG 2520
Qy 841 GlyArgArgLeuThrPheSerSerProLeuProArgCyAspAlaGluCyAspGlyAsnSer 860
Db 2521 GCGAGAGGCGGCTCTGAGACTCCCTCTTCCAGAGTGTCTTCTGAGTGTGGAAATTC 2580
Qy 861 ValThrGlyThrGlnGlyThrLeuLeuSerProAspPheProValAAsnTyAsnAsnAsn 880
Db 2581 GTCAAGAGCACTAGGAGTACTTGTCTGCCCAACTTCTCTGAGACTCAATTAACAAT 2640
Qy 881 HisGluCysHisLeuTySerHisLeuGlnThrGlnProGlyValGlyHisLeuValAlaArg 900
Db 2641 CATGAATGACATCACTCAATCCAGACCCAGGAGGAGGAATTCAGCTGAAAGCCAGG 2700
Qy 901 AlaPheGluLeuSerGluGlyAspValLeuValTyAspGlyAsnAsnAsnSerAla 920
Db 2701 GCATTGCAATCTTCCAGAGAGATGTCTCAAGTTTATATGACCAACCACTCCGCC 2760
Qy 921 ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAsnSerTherSer 940
Db 2761 CGTTGCTGGAGTTTATACCACTTGTGAGATGAGGGGTGACTTGAACAGACATCC 2820
Qy 941 SerSerLeuThrPheLeuAspPheHisLeuThrAspAlaGluAsnTherSerGlyPheGluLeu 960
Db 2821 AGCAGTGTGGCTGATTCATCACTGATGCTGAAGAACCAAGCAAGGAGCTTGAAGCTG 2880
Qy 961 HisPheSerSerPheGluLeuHisLeuGlyAspProGlyThrProLeuPheGlyTTrp 980
Db 2881 CACTTTTCCAGCTTGAACATCAATATGAGAGACCCAGAAACCCCAAGTTGGCTAC 2940
Qy 981 LysValHisAspGluGlyHisPheHisGlySerSerValSerPheSerCyAspProGly 1000
Db 2941 AAGGTTATGATGAGATGATTTTTCAGGAGAGCTCGGTCTTCACTGTGACCTGGA 3000
Qy 1001 TyTrSerLeuArgGlySerGluLeuLeuCysLeuSerGlyValAlaGlyGlyThrTTrpAsp 1020
Db 3001 TACAGCTTGGGGGTATGAGAGCTGTGTGTGAGTGAAGCCCGGAGCTTGGAGC 3060
Qy 1021 ArgProLeuProThrCysValAlaGluCyAspGlyValThrValArgGlyGluValSerGly 1040
Db 3061 CGGCTTGGCCCACTGTGTCCCGAGTGTGGAGGAGCACTGAGAGAGAGTGTCCGGG 3120
Qy 1041 GlnValLeuSerProGlyTTrpProAlaProTyGluHisAsnLeuAsnCyHisLeuThrPhe 1060
Db 3121 CAGGTGTGTACCCGGGATTCAGCTCTTATGAAACAACTCAATGCTCATCTGGAGC 3180
```

QY 1061 IIEGLAAGLALAGLYCYETHRIIEGLYLEUHIAPHELEUVALPHEASPTHICLUGLU 1080  
DB 3181 ATCGAAGCAGAGGCGGCTGACACATGGGCTTACACTTCGTTGGTTTGGACACAGAGAG 3240  
QY 1081 VALHIAPASVALLEUARGIIETTPASPGLYPROVALGLUSERGLYVALLEULEUVALG 1100  
DB 3241 GTTTCACGAGCTGCTGGGCACTGGGATGGGCTGTGGAGAGGGGGTCTGCTGAAGAG 3300  
QY 1101 LEUSERGLYPROVALLEUPROLYASAPLEUHISSERTHPHEASERVALLEULEU 1120  
DB 3301 CTGAGTGAGCCGCGCCCTGCCAAGGACCTGCATACACCTTCACTCGGTGCTCGAG 3360  
QY 1121 PHEASERTHPASPHETHERTHSELYSGINGLYPHEALAIIEGLINPHESEVALSER 1140  
DB 3361 TTCACACTGACTTCTTACACGACGAGAGGGCTTGCATTCATTTTCAGTCTCCACA 3420  
QY 1141 ALATHSERCYASAPASPPROGLYTLPROGLINANGLYSERATRGSERGLYASPSERTP 1160  
DB 3421 GCNAGTCTCTCAAGACCTGGGATCCGCAAGATGGAGTCCGAGTGGTGAAGTTGG 3480  
QY 1161 GLUALAGLYASPSERTHVALPHEGLINCYASAPPROGLYTYRALALEUINGLYSER 1180  
DB 3481 GAAGCCGCGACTCCACAGTGTTCAGTGTGACCTGGCTACGCGCTGCAGGGAGATGCA 3540  
QY 1181 GLUILESERCYVALLYVALIEGLUASAPRPHETHRPGINPROSERPROTHCY 1200  
DB 3541 GAGATCAGCTGTGCAAGATGAGAAACAGGTTCTTGGGAGCCGACCCGCAACATGC 3600  
QY 1201 ILEALAPROCYSGIYGLYASPLEUETHGLYPROSERGLYVALILEUSERTPROANTY 1220  
DB 3601 ATCCCTCCCTCGGGGGAGACCTACAGACCACTTGAGTCACTCCCTCAACAAATTC 3660  
QY 1221 PROGLUPROTYRPROPROGLYLYSGIUCYASPTTPHYSEVALTHYVALSERPROAP 1240  
DB 3661 CCAGAACCTTACCCGCGCAGGCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3720  
QY 1241 VALILEALEUVALPHEASERLEPHEASLEUENGLUPROGLYTYRAPHLEUHI 1260  
DB 3721 GTCAATCCCTCGGATTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACT 3780  
QY 1261 TYRAPHGLYARGASERTLEUSERTPROLEUIIEGLYSEPHETRYGLYSEGLINLEU 1280  
DB 3781 TAGACGAGAGGAGCTCTCAAGCCCTCATAGGAAGCTTTATAGGCTCCAGCTCCA 3840  
QY 1281 GLYARGILEGLUSERSERSEASERLEUENGLIAPHEARGSERAPALASER 1300  
DB 3841 GGCAGCATTTGAAAGCAGACCAACAGCTCTCTCGCTTCGCGAGCGATGATCTGG 3900  
QY 1301 SERASAPALAGLYPHEVALILEAPTYRTHGLUASPROARGGLUSERCYSPHEAS 1320  
DB 3901 AGCAATGCTGGCTTCTGCTATTTGCTTACAGAAACCCGCGGAGTCATGTTTAACT 3960  
QY 1321 GLYSERTILEYASANGLYTHARGVALIGLYSERASPLEULEUENGLYSESERVAL 1340  
DB 3961 GGTTTCATCAAGAACGAGCACACGCGTGGGCTCGACTGAAGTGGGCTCTCCCTCAC 4020  
QY 1341 TYRTHCYAHIEGLYGLYTYRGLIUVAGIUGLYTHSETHLEUSERTCYSTLEU 1360  
DB 4021 TACTACTGCCACGGGGGCTACAGAGTGGAGGACTTCGACCTTGAAGCTGCTGGGG 4080  
QY 1361 PROASPGLYLYASPROVALTRPASNANPROARGPROVALCYSTHVALPROCYSGI 1380  
DB 4081 CCTGATGGGAGGCGCTGTGGAAACATCCCGGCGAGTGTGCAACGCCCCCTGTGGGGA 4140  
QY 1381 GINTYVALIGLYSERAPGLYVALIALLUSERTPROANTYRPROGLINANTYRTHSER 1400  
DB 4141 CAGTATGTTGGGTTCCGACGAGTGTGCTTGTCCCACTACCCCAAGACTACAGCACT 4200  
QY 1401 GLYGLINILECYALEUETHYRPHETHYRVALPROLYASAPTYRVALPHEGLINPHE 1420  
DB 4201 GGAAGATCTGCTTTTATTCTTACTGTGCCCCAAGACTATGTGTGTGGCCAGTTTC 4260

QY 1421 ALAPHEPHEHITHRALALEUASAPVALIAGIUVALHIAPASGLYHISSEGLINHIS 1440  
DB 4261 GCCTTCTTTCACAGGCGCCCTCAACAGCGTGTGGAGTTCAGAGCGCCACAGCCAGC 4320  
QY 1441 SERARLEULEUSERTSERLEUSERTGLYSETHITHRGLYGLUSERLEUPROLEUAL 1460  
DB 4321 TCGGGCTCTCAAGCTCCCTCTCGGGCTCCATACAGGAGAAATCACTGCTTGGCCACC 4380  
QY 1461 SERASAPVALILEUIIEYSPHEASERVALYSGIYLEUVALPROVALARGIYPHEHIS 1480  
DB 4381 TCCATTCAGATTCTCATTTAGTTACGCGCAAGGCTTGACACAGCCAGAGGCTTCAC 4440  
QY 1481 PHEVALTYRGLINALLAVALPROARGTHSERALATHRGINCYSESERVALPROGLUP 1500  
DB 4441 TTTGTTCACCAAGCGGTTCTCGAACCAAGCCGACAGAGTGCAGTGTGCGGAACC 4500  
QY 1501 ARGTYRGLYLYASRLEUENGLYSERAPPHESERTVALIAGIUVALLIIVALARGPHEGL 1520  
DB 4501 CCTATGCAAGAGGCTGGGCAAGTACCTTCTGGTGGGGCCATGCTCGCTTGGAAATGC 4560  
QY 1521 AASERGLYTYRALALEUENGLYSEPROGLUIIEGLUCYLEUPROVALPROGLYALA 1540  
DB 4561 AACTCCGCTATGCTCTGAGAGGCTGCGCAGAGATCGATGCTCTCTGCTGGGCC 4620  
QY 1541 LEUALAGINTROPASVALISERVALPROTHRCYSEVALIAPROCYSGIYGLYASPLEU 1560  
DB 4621 TTGGCCCAATGGAATGCTCAGCGCCACGTTGTGGTGGCGGTGTGAGGCAACTCACA 4680  
QY 1561 GLUARGARGLYTHRIIELEUSERTPROGLYPHETPROGLUPROTYRLEUASERLEU 1580  
DB 4681 GAGCGCAGGGGCAACATCTGCTGCTGCTTCCAGAGCCGTACCTCAACAGCTCAAC 4740  
QY 1581 CYSEVALITRPHYSEIIEVALIAPROGLIYVALAGLYIIEGLINILEUVALI 1600  
DB 4741 TGTGTGTGAAGATGTGTGTGTCCGAAAGCCCTGCGATCCAGATCCAACTTGCAGTTT 4800  
QY 1601 VALTHRGINGLINANTTRPASPSELEUENGLIUALPHEASPGIYALAPASANTH 1620  
DB 4801 GTGACAGAGCAAGACTGGGACTGCTGAGAGTATTTATGTTGAGATACACTGTAC 4860  
QY 1621 METLEUENGLYSETPHESEGLYTHRTHYVALPROALALEUENGLIANSERTHSEAS 1640  
DB 4861 ATGCTGGGAGATTTCTAGGAAACAGCTGCTGCTCTTGAAACAGACCTCCACAG 4920  
QY 1641 LEUTYRLEUHIAPHEYRYSERASPILESERVALSERIALIAGLYPHEHISLEU 1660  
DB 4921 CTCTACTCTTCTTCTACTGATATCAGATATCAGCTGCTGCTTCCACTTGGAGTAC 4980  
QY 1661 LYETHRYVALIGLYLEUSERTCYEPROGLUPROVALIAPROSERASANGLYVALI 1680  
DB 4981 AAAACGGTGGGCTGAGCAGATTGCCGAACCTGTGTGCTGCAGTAAAGGGGTGAGACT 5040  
QY 1681 GLYGLUARGTYRLEUVALIAPASAPVALISERPHENGLINCYSGIUPROGLYTYRAL 1700  
DB 5041 GGCAGAGCTTACTGTGGAATGATGTGTGTCTTCCAGTGTAGCGCGGATATGCTTC 5100  
QY 1701 GINGLYHIALIHISILESECYSEMEPROGLYTHYVALARGATTPASANTYRPRO 1720  
DB 5101 CAGGGCAGCCACCAATCTCTCTCATGCGGAAAGTGGGAGATGAACTACCTCTCT 5160  
QY 1721 PROLEUCYSTILEALAGINCYSGIYGLYTHYVALIGIUGIUMEGIUGIYVALI 1740  
DB 5161 CCACTGTGATTCACAGTGTGTGGGGAACAGTGAAGAGATGAGGGGGTGAATCTCGAG 5220  
QY 1741 PROGLYPHEPROGLYASANTYRPROSERASMEASPCYSESERTRPULYIIEALALEU 1760  
DB 5221 CCGGCTTCCAGCAACTTACCCCAATTAATGATGAGTGTGCTCTGGAATAATACAGCTCC 5280  
QY 1761 VALIGLYPHEGLYVALIHSIIEGLINPHELEUASPNHESERTHGLUPROASNHISAP 1780  
DB 5281 GTGGGCTTGGAGACTCAATCCAGTTCTGAACTTCTCAACGAGGCCCAACAGACTAC 5340  
QY 1781 IIEGLUIIEARGASANDLYPROTYRGLIUTHIRSERARGMECMEGLYARGPHESE 1800

```
Db 5341 ATGAAATCCGGAATGCCCCCTATGAGACAGCCGATGAGGAAATTCAGTGAAGC 5400
Oy 1801 GLeuEUPrOSerSerLeuLeuSerThrsEriSeglUThrThValTYrPheHisSerasp 1820
Db 5401 GAGCTTCACAGCTCCCTCTCCACGTCACGAGACACCGGTATATTTCCACAGCCAC 5460
Oy 1821 HisSerGlnAAsnArgProGlyPheIysLeuGluTYrGlnAlaTYrGlnLeuGlnIuys 1840
Db 5461 CACTCCCAAGATCGGCCAGATTCAGCTGAGATTCAGGCTTATGAATTCACAGAGTGC 5520
Oy 1841 ProAspProGluProPheAlaAsnGlyIleValArgIValAGlyTYrAsnValGlyGln 1860
Db 5521 CCAGACCCAGAGCCCTTTGCCAATGGCATTTGAGAGGAGCGCTGCTACAGCTGAGACA 5580
Oy 1861 SerValThrPheGluCyLeuLeuProGlyTYrGlnLeuThrGlyHisProValLeuThrCys 1880
Db 5581 TCAGTGACCTTCGATGCTCCCTCCGGGGTATCAATTGACCTGGCCACCTGTCTCTCACTGT 5640
Oy 1881 GlnHisGlyThrAsnArgAsnTrpAspHisProLeuProIysCysGluValProCysGly 1900
Db 5641 CAACATGGCACCAACCCGAACTGGGACCAACCCCTGCGCAAGGTGAAGTCCCTGTGGC 5700
Oy 1901 GlyAsnIleThrsSerSerAsnGlyThrValTYrSerProGlyPheProSerProTYrSer 1920
Db 5701 GGGAACATCATCTTCCACAGGCACTGTACTCCCGGGGTTCCAGCCCTGACTCC 5760
Oy 1921 SerSerGlnAspCysValITrpleuIleThrValProIIleGlyHisGlyValArgLeuAsn 1940
Db 5761 AGCTCCCGAGACTGTGTGGCTGATCAACCGGCCATTGGCCATGGCGGTCCGCTCAAC 5820
Oy 1941 LeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTrpAspGlyProGln 1960
Db 5821 CTCAGCTCTGTCAGACAGAGCCCTCTGAGATTTCACTACCATTTGGGATGGGCCACAG 5880
Oy 1961 GlnThralaProArgLeuGlyValPheThrArgSerMetAlaIlyIysThrValGlnSer 1980
Db 5881 CAACAGACACACGGCTCGGGCTTTCACCCGGACATGGCCAGAAACAGTGCAGGT 5940
Oy 1981 SerSerAsnGlnValLeuLeuIysPheHisValArgAspAlaAlaThrGlyGlyIlePheAla 2000
Db 5941 TCATTCACACAGGTCTCTCTCAAGTTCACCGGTGATGACGACACAGGGGGATCTTCCCC 6000
Oy 2001 IleAlaPheSerAlaTYrProLeuThrIysCysProProThrIleLeuProAsnAla 2020
Db 6001 ATAGCTTCTCGCTTATCTACCTACCAAAATGCCCTCTCCACCATCTCCCAAGCC 6060
Oy 2021 GluValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTYrArgCysLeu 2040
Db 6061 GAAGTCGTACAGAGAAATGAAGAATTCATATAGTGAATCGTACGCTACAGATGCTCT 6120
Oy 2041 ProGlyPheThrLeuValGlyAsnGluIleLeuThrCysIysLeuGlyTYrThryLeuGln 2060
Db 6121 CCTGGCTTACCTTAACTAGTGGGGAATGAAATTTCTGACCTGCAAACTTGGAACCTACCTG 6180
Oy 2061 PheGluGlyProProProIleCysGluValHisCysProThrAsnGluLeuLeuThrsP 2080
Db 6181 TTGAGGAGACACCCCGATATGTGAATGACGTCTCAACAACAAAGACCTTCTGACACAC 6240
Oy 2081 SerThrGlyValIleLeuSerGlnSerTYrProGlySerTYrProGlnPheGlnThrCys 2100
Db 6241 TCCACAGGCGTGAATCTGAAGCAGAGCTACCTCGAAGCTATCCCAAGTTCCACAGCTGC 6300
Oy 2101 SerTrpleuValArgValGluProAspTYrAsnIleSerLeuThrValGluTYrPheLeu 2120
Db 6301 TCTTGCGTGTGAGAGTGGAGCCGACATATAACATCTCCCTCACAGTGAAGTACTTCTC 6360
Oy 2121 SerGluysGlnTYrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProIeu 2140
Db 6361 AGCGAAGAACATATGATGAGATTGATTTTGTGTGATTCATCGGACAGCAAGTCTCTG 6420
Oy 2141 LeuIysAlaLeuSerGlyAsnTYrSerAlaProLeuIleValThrsSerSerAsnSer 2160
Db 6421 CTGAAGCCCTCAGTGGGAATTACTACGCTCCCTGATTTGTCACACGCTCAAGAACTCT 6480
Oy 2161 ValTYrLeuAspArgTrpSerSerAspHisAlaIlyTrpAsnArgIysGlyPheIysIleArgTYr 2180
Db 6481 GTGTACCTGCGTGTGTATCTGATCAGCCCTACAAATGGAAAGGGCTTCAAGATCCGCTAT 6540
Oy 2181 SerAlaProTYrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr 2200
Db 6541 TCAGCCCTTATCTGACGCTCCGACAGGCTCCATGAGCTTCATCCATGGCTTCACTTACGCAAGCC 6600
Oy 2201 SerThrGlnProGlyIysSerIleHisPheGlyCysAsnAlaGlyTYrArgLeuValGly 2220
Db 6601 AGCACCCAGCCCGGGGCTCCATCCACTTTGGCTGCAACGCGGCTACCGCTGTGGGA 6660
Oy 2221 HisSerMetAlaIleCysEThrArgHisProGlnIlyTYrHisLeuTrpSerGlnAlaIle 2240
Db 6661 CACAGCATGGCCATCTGTACCCGGCACCCCGAGGATCAACACGTGTGAGGAAGCAATC 6720
Oy 2241 ProLeuCyseGlnAlaLeuSerCysGlyLeuProGluAlaProIysAsnGlyMetValPhe 2260
Db 6721 CCTCTCTGTACAGCTCTTCTCTGTGGCTTCTGAGGCCCCCAAGAAATGAAATGCTGTT 6780
Oy 2261 GlyIysGluTYrThrValGlyThrIysAlaValTYrSerCysSerGluGlyTYrHisLeu 2280
Db 6781 GGCAAGAGATACACAGTGGGAACCAAGCCGTGTACAGCTGACATGAAGGCTACCACTC 6840
Oy 2281 GlnAlaGlyAlaGluAlaThrAlaGluCyLeuAspPheArgIleuTrpSerAsnArgAsn 2300
Db 6841 CAGCAGGCGCTGAGGCCACAGCAGAGTGTGACACAGGCCATATGAGAACACCGCAAT 6900
Oy 2301 ValProProGlnCysValProValThrCysProAspValSerSerIleSerValGlnHis 2320
Db 6901 GTCCACACAGTGTGCTCTGTGATCTTGTCTGTATGTCACTAGACATCAGCTGAGCAT 6960
Oy 2321 GlyArgTrpArgLeuIlePheGluThrGlnIlyTrpGlnPheGlnAlaGlnLeuMetLeuIle 2340
Db 6961 GGCCGATGAGAGCTTATCTTTGAGACACAGATACAGTTCCAGGCCAGCTGATCTCATC 7020
Oy 2341 CysAspProGlyTYrTYrTYrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyIys 2360
Db 7021 TGTACCTCGCTACTACTATACAGGCCAAAGGTACTCCGCTGACGCCAAATGGCAAA 7080
Oy 2361 TrpSerLeuGlyAspSerTrpProThrCysArgIleIleSerCysGlyGluLeuProIle 2380
Db 7081 TGGAGCTCGGGAGCTCTAGCCCACTGCGCAATCATCTCTGTGAGAGCTCCGAT 7140
Oy 2381 ProProAsnGlyHisArgIleGlyThrLeuSerValTYrGlyAlaThrAlaIlePheSer 2400
Db 7141 CCCCCAATGGCCACCGCATCGGAACACTGTCTGTACGGGGCAACAGGCATCTTCTCC 7200
Oy 2401 CysAsnSerGlyTYrThrLeuValGlySerArgValArgIleCysMetAlaAsnGlyLeu 2420
Db 7201 TGCATATTCGGATACACACTGTGTGGCTCCAGGGTGTGTGATGTCATGGCCAAATGGGCTC 7260
Oy 2421 TrpSerGlySerGluValArgCysLeuAlaGlyHisCysGlyThrProGluProIleVal 2440
Db 7261 TGGAGTGTCTGAAGTCCGTGCTGCTTGTGGAACCTGTGGACTCTCTGACCCCATTTGC 7320
Oy 2441 AsnGlyHisIleAsnGlyGluAsnTYrSerTYrArgIlySerValIlyTYrGlnCysAsn 2460
Db 7321 AACGGAACATCATATGGGGAATACAGCTACCGGGGAGTGTGTGATACCAATGCAT 7380
Oy 2461 AlaGlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisIleTrpSer 2480
Db 7381 GCTGCTTCCGCGCTGATCGCATGTCTGTGGCATCTGCGACAGCAGATCATCTACGTGTG 7440
Oy 2481 GlyIysEThrProPheCysValProIleThrCysGlyIlyHisProGlyAsnProValAsnGly 2500
Db 7441 GGCAGAGACCCCTTCTGTGTGTCATATACCTGTGGAACCCAGGCAACCTGTATAAGGC 7500
Oy 2501 LeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValIysPheValCysAsnProGly 2520
Db 7501 CTCACCTCAGGGTAAACAGTTTAACCTCAAGATGTGTCAAGTTGTTTGAACCTGTGGG 7560
```

|    |      |  |      |
|----|------|--|------|
| QY | 2521 | TYRMEALAGLGLVALLAALARGSERGINCYSEUVALASERGLYGLNTPSRASPMEC     | 2540 |
| DB | 7561 | TTATATGCTGAGGGGGCTCTAGTCCCAATCCCTGGCCAGGGGCAATGAGTGAATG      | 7620 |
| QY | 2541 | LEUPROTHRCYARGLLEILEANCYSTRHARSPROGLYHISGLNGLUASERVALARG     | 2560 |
| DB | 7621 | CTGGCCACCTCGAGATCATCACTGTACAGATCCTGGACACCAAGAAAATAGTTCCT     | 7680 |
| QY | 2561 | GLNVALHISALASERGLYPROTHARPHESERPLEGLYTHRVALLSERTRYARGCYA     | 2580 |
| DB | 7681 | CAGGTTCACAGCGGGGGCCGACAGGTTCAGCTTCGGCACCACTGTCTTACCGGTGC     | 7740 |
| QY | 2581 | ASNHISGLYPHETRYLEULEUGLYTHRVALLSERCYSGNLGLYASPGLYTHTRP       | 2600 |
| DB | 7741 | AACCAAGGCTTCTACTCTCTGGGACCCCACTGCTCAGCTCCAGAGATGSCATGG       | 7800 |
| QY | 2601 | ASPARPROARGPROGLNLYSEUVALSERCYSGLYHISPROGLYSEPROTHIS         | 2620 |
| DB | 7801 | GACCGTCCCGCCCGCAGTGTCTTGGGTCTTGGGCATCCGGGCTCCCGCTCAC         | 7860 |
| QY | 2621 | SERGLMETSERGLYASPSERTYTHRVALLVALVALARGTRYSERCYGLLEGLY        | 2640 |
| DB | 7861 | TCCCAATGTCTGAGACATTAATCTGSGAGCAGTGGCGTACAGCTGATCGCGC         | 7920 |
| QY | 2641 | LYSARGTHRVALLGLYASNSETRHARGMETCYSGLYLEUAPRGLYHISSTPTTHGLY    | 2660 |
| DB | 7921 | AAGGTAATCTCGTGGGAAACAGCACCCGATGTGGGCTGATGACATGACATGAC        | 7980 |
| QY | 2661 | SERLEUPROHISCYSESERGLYTHSERVALGLYVALCYSGLYASPPROGLYILEPROLA  | 2680 |
| DB | 7981 | TCCCTCCCTCACTCTCAAGAACAGCGTGGAAATTTGGCGTATCCCTGGGAATCCGGCT   | 8040 |
| QY | 2681 | HISGLYILEARGLEUGLYASPSERPHASPPROGLYTHRVALLMETARGPHESERCYGLN  | 2700 |
| DB | 8041 | CATGGCATCCGCTTGGGGGACAGCTTTGATCCAGCAGCTGTGATGCGCTTCACTGTGA   | 8100 |
| QY | 2701 | ALAGLYHISVALLEUARGGLYSESERCYGLNARGTHRCYSGNLAAANGLYSETRPSE    | 2720 |
| DB | 8101 | GCTGGCCACGTCCTCCGGGATCGTCAAGCGCACCTGTCACCAATGCTGCTGAGC       | 8160 |
| QY | 2721 | GLYSEGINPROGLUCYSGLYVALILLESERCYSGLYASPPROGLYTHRVPROSERANALA | 2740 |
| DB | 8161 | GGCTCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG     | 8220 |
| QY | 2741 | ARGVALVALPHESEARPGLYLEUVALPHESESESERILEVALTYRGLUCYARGGLN     | 2760 |
| DB | 8221 | CGAGTGTGTTCAGATGATGCTGCTGCTTCTCCAGCTCTATGATGCTCGGGA          | 8280 |
| QY | 2761 | GLYTRYTRYLATHRCGLYLEULEUSERARGHISCYSESERVALAANGLYTHTRPTHRG   | 2780 |
| DB | 8281 | GGATCTACGCGCACAGGCTGCTCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG      | 8340 |
| QY | 2781 | SEARSPROGLUCYSEUVALILEANCYSGLYASPPROGLYILEPROLAAANGLYLEU     | 2800 |
| DB | 8341 | AGTGACCTCGAGTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG        | 8400 |
| QY | 2801 | ARGLEUGLYASNPARGHARGTRYASNLSTHVALTHTRYGLNLYSEUVALPROGLYTYR   | 2820 |
| DB | 8401 | CGGCTGGGCAATGATCTGATGATGATGATGATGATGATGATGATGATGATGATG       | 8460 |
| QY | 2821 | METMETGLYSEHISARGVALSERVALLEUSERCYSTHLYASPARGLYTHRVANGLY     | 2840 |
| DB | 8461 | ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG       | 8520 |
| QY | 2841 | THRVSPROVALCYSELYSALALEUWETCYALSPROPROLEULIETPROANGLYLYS     | 2860 |
| DB | 8521 | ACCAAGCCCGCTCGCAAGCTCTCATGTGCAAGCCACCTCCGCTATCCCAATGGAG      | 8580 |
| QY | 2861 | VALVALGLYSEARSPHEMETRPGLYSESERVALTHTRYALAICYLEUGLGLYTYR      | 2880 |
| DB | 8581 | GTGGTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 8640 |

  

|    |      |  |      |
|----|------|--|------|
| QY | 2881 | GLNLEUSERLEUVALALPHEHRCYSGLYVALANGLYSETRPTHRGGLYLEU          | 2900 |
| DB | 8641 | CAGCTCTCCCTGCGCCGCGGTGTTCACTGTGAGGGAATGGGCTCGAGCCGAGAGCTG    | 8700 |
| QY | 2901 | PROGLNLYCYPHEPROVALPHECYSGLYASPPROGLYVALPROSERARGLYARGARGGLN | 2920 |
| DB | 8701 | CCTCAATGTTTCCCTGATGTTCTGGGGGATCTGTGTCTCCCTCCCTGGGAGAGAG      | 8760 |
| QY | 2921 | ASPARGLYPHESETRYARGSERSEVALSERPHESECYAHISPROPROLEUVALLEU     | 2940 |
| DB | 8761 | GACCGAGGCTTCTCTCAAGATCATCTGCTCTCTCTGCAATCCCTCTGATGCTG        | 8820 |
| QY | 2941 | VALGLYSEPROARGHARGPHECYSGLYSEARPGLYTHTRIPSEGLYTHRGINPROSE    | 2960 |
| DB | 8821 | GTGGGCTTCCACGAGGTTTGGCCAGTCAATGAGCATGAGATGAGACCCAGCCAGC      | 8880 |
| QY | 2961 | CYSLASPPROTHRVALLSERCYSLAASPPROGLYVALPROGLNPHRGGLYILEGLN     | 2980 |
| DB | 8881 | TGCATGATCCGACCTTACACAGTGTGGAGACCTGTGTGTGCAAGTTGGATGACG       | 8940 |
| QY | 2981 | ASNAASERGLNGLYTRYGLNVALGLYSETRHVALLEUPHARGCYGLNLYSGLYTYR     | 3000 |
| DB | 8941 | AACAAATTCAGGGCTACAGGTTGGAAGCACAGTCTTCCCTTCCCAAAAAGCTAC       | 9000 |
| QY | 3001 | LEULEUGLNGLYSETRHTRYARGHRCYSLAUPROANLEUTHTRIPSEGLYTHRVPRO    | 3020 |
| DB | 9001 | CTGCTTCAAGGCTTCCACACAGACCTGCTCCCAACCTGACCTGAGTGAACCCCA       | 9060 |
| QY | 3021 | PROAERCYVALPROHISHISCYAARGLINPROGLNTHRVPROTHHISALAASVALGLY   | 3040 |
| DB | 9061 | CCTGACTGTGTCCCGACCACTGACAGGAGCAGAGAGCCAGAGCATGCGAGGCTG       | 9120 |
| QY | 3041 | ALALEUAPLEUPROSEMERGLYTRYTHRVALLTHRVALLAARGARGALASERPRO      | 3060 |
| DB | 9121 | GCCCTGATTTGCCCTCATGAGGCTACACGCTCTTACTCTGCGACGAGGCTTCTCC      | 9180 |
| QY | 3061 | SEARGLVALAIPROSETRHVALPROVALARGHRCYSLAAGLYVALASERARG         | 3080 |
| DB | 9181 | TCAAGGAGGCTTCCGAGCAGCAGCTGCAAGGAGGAGTGGAGTGAAGGCAAGCCG       | 9240 |
| QY | 3081 | PROSERALATPARSERGLYPROVALGLYASPPROSETRHLEUPROGLYSEHISARG     | 3100 |
| DB | 9241 | CCATCTGCTGAGATGCTCGGCGCAGTGGAGACCACTCAACATCGCCGGAGCCACCC     | 9300 |
| QY | 3101 | SEPROLYSPRO 3104   |      |
| DB | 9301 | TCACCAAGCCT 9312   |      |

  

RESULT 2  
 ADH71165  
 ID ADH71165 standard; DNA; 10136 BP.  
 XX  
 AC ADH71165;  
 XX  
 DT 25-MAR-2004 (first entry)  
 XX  
 DE Human gene of the invention NOVAP SEQ ID NO:61.  
 KW ds; gene; human; cytostatic; immunomodulator; neuroprotective; nootropic;  
 KW anorectic; antidiabetic; antimicrobial; antihypertensive; gene therapy;  
 KW obesity; diabetes; infectious disease; metabolic syndrome X;  
 KW dyslipidaemia.  
 OS Homo sapiens.  
 XX  
 PN W02003102155-A2.  
 XX  
 PD 11-DEC-2003.  
 XX  
 PF 03-JUN-2003; 2003MO-US017430.  
 XX  
 PR 03-JUN-2002; 2002US-0385120P.

[illegible]

```
Db 301 AAGAGGGCTATACCTGCAAGGCTCAAGCGGATCACTGTATGAATGAGCAATG 360
Qy 121 PheAlaAlATpSerAspHisArgProValCyArgAlaArgMetCysAspAlaHisLeu 140
Db 361 TTTCGGCGCTGGAGGACCAAGGCGAGTCTGGCCGAGCCGATGTGTATGCCACTT 420
Qy 141 ArgGlyProSerGlyLeileIleThrSerProAnPheProileGlnTyAspAsnAla 160
Db 421 CGAGGCCCTTGGGACATCACTCCCAATTTCCCAATTCAGATGACAAACATATCA 480
Qy 161 HisCysValTrrIleIleThrAlaLeuAnProSerIleValIleLeuLeuAlaPheGlu 180
Db 481 CACTGTGTGTGATCATCAAGACCTCAACCCCTCAAGGTATCAACTGACTGCTTGG 540
Qy 181 GluPheAspLeuGluArgGlyTyArgPheLeuThrValGlyAspGlyGlyGlnAspGly 200
Db 541 GAGTTGATTTGGAAGGGGCTATGACACCTGACGCTGGGTGTGTGTGTGTGTGTGTGT 600
Qy 201 AspGlnIleThrValIleuTyIleSerGlnAsnAlaCysSerAspSerProHisThrPro 220
Db 601 GACCAAGAACAGCTTCTTACATGTCTCAAAATGCTTGACGTGACAGCCCTCACACCCCA 660
Qy 221 GlySerArgIleProGluSerMetSerGlyAspIleTrrArgGlnIleTyTrpThrValLeu 240
Db 661 GGCTCTCCATCCAGAGAGCATGTCTGGGACATCTGGAGCATGTGAGCAATGACCTGACTT 720
Qy 241 GluIleCysArgAspIleSerSerSerAspAlaArgSerGlySerValAlaGlySerPro 260
Db 721 GAGATCTCTGTGACATTAAGCATTAAGCATTAAGCATTAAGCATTAAGCATTAAGCATTA 780
Qy 261 LysThrSerAsnAlaValGluLeuValAlaProGlyThrGluIleGlnGlnIleGlySerCys 280
Db 781 AAGCTTCTTAATGCTGTGAGCACTTGCTGCTGTGGACAGAGANTCGAGCGGCACTTGC 840
Qy 281 GlyAspProGlyIleProAlaTrrGlyArgArgGluGlySerArgPheHisIleGlyAsp 300
Db 841 GGTACCTTGGCATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Qy 301 ThrLeuIlePheGluCysGlnProAlaPheGluLeuValGlyGlnIleValAlaIleThrCys 320
Db 901 ACACCTCAAGTTGAGTGCAGCCGCTTTGAGTGTGGAGAGAGAGAGAGAGAGAGAGAGAG 960
Qy 321 GlnIleAsnAsnGlnIlePheSerAlaValLeuPheProGlyCysValPheSerCysPhePheAsn 340
Db 961 CAAAGATATACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
Qy 341 PheThrSerProSerGlyValIleuSerProAnTyrrProGluAspTyrrGlyAsnHis 360
Db 1021 TTTCACGAGCCGCTGTGGGTTGTCTGTCTTCCCACTACCAAGAGATATGAGCAACAC 1080
Qy 361 LeuHisCysValTrrPheLeuIleLeuAlaArgProGluSerArgIleHisLeuAlaPheAsn 380
Db 1081 CTCCACGTGTCTGTGCTCATCTGCGCAGGCTGAGAGCGGATCCACTGCGCTTCAAC 1140
Qy 381 AspIleAspValGluProGlnPheAspPheLeuValIleLeuAspGlyAlaThrAlaGlu 400
Db 1141 GACATTCAGTGAAGCTCACTGTTGATTTCTGTGTATCAAGATGAGGCGCACCCCGGAG 1200
Qy 401 AlaProValIleuGlyThrPheSerGlyAsnGlnIleuProSerSerIleThrSerSerGly 420
Db 1201 GCGCCGCTCTGGGACCTTCTCAGGAAACAGCTTCCCTCTCTCATCAAGCATGAGC 1260
Qy 421 HisValAlaArgLeuGluPheGlnThrAspHisSerThrGlyValArgGlyPheAsnIle 440
Db 1261 CAGGTGAGCCGCTTCAGATTCAGACTGACCACTCCACAGGAGAGAGGAGGCTTCAACATC 1320
Qy 441 ThrPheThrThrPheArgHisAsnGluCysProAspProGlyValProValAsnGlyIys 460
Db 1321 ACTTTTACCACTTCGACACAAACAGATGCTGGGATCTGGGCTTCAAGTAAATGGCAAA 1380
Qy 461 ArgPheGlyAspSerLeuGlnIleuGlySerSerIleSerPheLeuCysAspGlyGlyPhe 480
1381 CGGTTTGGGACAGCCTCCAGCTGGGAGCTCCATCTCTCTGTGTATGAAGCTTC 1440
Qy 481 LeuGlyThrGlnIleuGlyLeuTrrIleThrCysValIleuGlyGlySerValIleTrr 500
Db 1441 CTTCGAGCTTCAGGAGCTTCAGAGACATCACTGCTGTCTGTAAGAGAGGAGCGGTGCTGG 1500
Qy 501 AsnSerAlaValLeuArgCysGluAlaProCysGlyGlyHisLeuThrSerProSerGly 520
Db 1501 AACAGGCTGTGTGTGGGTGTGAGTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
Qy 521 ThrIleLeuSerProGlyTrrProGlyPheTyrrIleAspAlaLeuSerCysAlaTrrVal 540
Db 1561 ACCATCTCTCTCCGGCTGGCTGCTGCTTCAAGATGCTTGAAGTGTGTGTGTGTGTGTGT 1620
Qy 541 IleGluAlaGlyProGlyTyrrProIleValIleThrPheAspArgPheIleThrGluVal 560
Db 1621 ATTAGAGCCCAAGCAGGCTTACCCATCAAAATCACTTCAGATTCAGATTCAGATTCAG 1680
Qy 561 AsnTyrrAspThrLeuGluValArgAspGlyArgThrTyrrSerAlaProLeuIleGlyVal 580
Db 1681 AACTATGACACCTGGAAGTACGCGATGGGGGACTTACTAGCCCTTGATCGGGATT 1740
Qy 581 TyrrHisGlyThrGlnValProGlnPheLeuIleSerThrSerAsnTyrrIleuThrLeu 600
Db 1741 TACCAAGGAGCCAGAGTTCCCAATCTCTCATACAGACACAGCACTACCTCACTCTC 1800
Qy 601 PheSerThrAspIleSerHisSerAspIleGlyPheGlnLeuArgTyrrGlyThrIleThr 620
Db 1801 TTCTTACCAAGCAAAAGCATCTCGGACATCGGCTTCCAGCTCCGCTTATGAACATTAACA 1860
Qy 621 LeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGlnArgHisGlyAsn 640
Db 1861 CTGCAGTACAGCACTGTCTGTGATCCAGAAATCCAGTAAATGACAGCGTCAAGGAAAT 1920
Qy 641 AspPheTyrrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrrThrLeuSerAsp 660
Db 1921 GACTTCTTACGTGGGCGGCTGTGATCTTCAAGCTGTGATCTGAGGCTTCACTTAAGTAC 1980
Qy 661 GlyGluProLeuGluCysGluProAnPheGlnIleTrrSerArgAlaLeuProSerCysGlu 680
Db 1981 GGGAGGCTCTGGAGGTGAGCCCACTTCAGTGGAGCGGGCGCTGCCAGTGTGAA 2040
Qy 681 AlaLeuCysGlyGlyPheIleGlnIleGlySerSerGlyTyrrIleLeuSerProGlyPhePro 700
Db 2041 GCTCTGTGTGTGCTTCAATCAAGGCTCCAGTGGAGCAATCTTGTGTCCAGAGGTTCTCT 2100
Qy 701 AspPheTyrrProAnAsnLeuAsnCysThrTrrIleIleGluThrSerHisGlyIysGly 720
Db 2101 GACTTCAACCCCAACAATGAACTGCACTGGATTTATGAAACATCTTCATGCAAGGAT 2160
Qy 721 ValPhePheThrPheHisThrPheHisLeuGluSerGlyHisAspTyrrLeuLeuIleThr 740
Db 2161 GTGTCTTCACTTCCACACCTTCCACCTGGAAGAGGCGCATATCACTCTCATCACT 2220
Qy 741 GluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgIleuProAlaPro 760
Db 2221 GAGAGCGGACGCTTCAACCCAGCCCTGAGGACCTTAACGATCTGGAGCGCAGCTCCC 2280
Qy 761 IleSerAlaGlyLeuTyrrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780
Db 2281 ATACGCGCTGTGCTTATGCACTTCACTGCGCAAGTCCGCTTCACTCTGATTTCTCC 2340
Qy 781 MetSerTyrrGluGlyPheAsnIleThrPheSerGlyTyrrAspLeuGluProCysGlyGlu 800
Db 2341 ATGTCTATGAAAGATTCACATCACTTCTCAAGATGACATTCGAGGCTGTGGAG 2400
Qy 801 ProGluValProAlaTyrrSerIleArgIysGlyLeuGlnPheGlyValGlyAspThrLeu 820
Db 2401 CCGAGGCTCCAGACCTACAGCATCCGGAAGGCTTGCACTTGGGCTGGGCGACACCTTG 2460
Qy 821 ThrPheSerCysPheProGlyTyrrArgLeuGluGlyThrAlaArgIleThrCysLeuGly 840
Db 2461 ACCTTCTCTGCTTCCCGGGATCCGTGTGAGGGGACCGCGCATCACGTGCTGGGG 2520
```

QY 841 GLVARGARGLEUTRPSERSEPRORUARGCYVALAAGLUCYAGLYASNSER 660  
 DB 2521 GGCGAGCGCGCCCTGTGAGAGCTCGCTCTGCCAAGTGTGTGCTGAGATGGAGATTCA 2580  
 QY 861 VALTHRGJYTHRGJNGLYTHRLLEUSESERPROANPHEPROVALASNTYRASNAN 880  
 DB 2581 GTCAAGGCACTCAGGAGTACTTGTGTGCCCAACTTCTCTGTGAATACATACAT 2640  
 QY 881 HISEGLUCYVLIETYSERILEGINTHRCINPROGLYVSGLYILEGLILEULYVALARG 900  
 DB 2641 CATAAATGCACTCATCTCATCCAGACCAGGAGGAGGAAATTCACCTAAGGCCAGG 2700  
 QY 901 ALAPHEGLILEUSESERGLIYASPVALLILEULYVALTYRASPGLYASASNASERLA 920  
 DB 2701 GCATTCGAACTCTCCGAGAGAGATGCTCAAGGTTATGATGGACAAACAACATCCGCC 2760  
 QY 921 ARGLEULEGLIYVALPHESETHISESERGLIUMECMGLIYVALTHRLLEUSESERTHSER 940  
 DB 2761 CGTTTCCTGGGAGTTTATAGCCATTCTGAGATGANTGGGGGTGACTTTGACAGCAATCC 2820  
 QY 941 SERSERLEUTRPLEUASPHEIETHRASPALAGLUANTHRSELYSGLYPHEGLILEU 960  
 DB 2821 AGCAGTCTGTGGCTGATTTCATCACTGATGCTGAAAACACACAGCAAGGCTTTGAACTG 2880  
 QY 961 HISPHESESERPHEGLILEULILEYSEGLIUSAPPROGLYTHRPROLYSPHEGLIYTR 980  
 DB 2881 CACTTTTCAGCTTTGAACCTAATCAATGATGAGGCCAGGAAACCCCAACTTTGGCTAC 2940  
 QY 981 LYVALHISAPRGJLUGLYHISPHENAGLYSESERVALSERPHESECYASAPPROGLY 1000  
 DB 2941 AAGTTTCATGATGAAAGTCATTTTGCAGGAGCTCCGTCTCTTCACTGTGACCTTGA 3000  
 QY 1001 TYRSEULEUARGLYSERGLIULEULEUCYSEUSESERGLYIUAARGYTHRTTPASP 1020  
 DB 3001 TACAGCCCTGGCGGTAGAGAGAGAGCTGTGTCTGAGTGAAGAGCCGCGACCTGGGAC 3060  
 QY 1021 ARGPROLEUPROTHRCYVALAAGLUCYAGLYVALTHRVALLARGGLYGLIUNVALSERGLY 1040  
 DB 3061 CGGCTCTGGCCACCTGTGTGCGCCAGTGTGAGAGGACAGTGAAGGAGTGTCCGGA 3120  
 QY 1041 GLINVALLEUSESERPROGLYTYRPROLAPROTYRGLIHISANLEUASCYILETRTHR 1060  
 DB 3121 CAGGTGCTGTCAACCGGATATCCAGCTCCCTATGAAACAAATCTCACTGATCTGACC 3180  
 QY 1061 ILEGUVALAGLUALAGLYCYETHRILEGLYLEUHIAPHELEUVALPHEASPTRHGLIUN 1080  
 DB 3181 ATCGAAGCAGAGCGCGCTGCACCATTTGGCTACACTTCTGTGTGACACAGAGAG 3240  
 QY 1081 VALHISAPVALLEUARGILETRPASPGLYPROVALGLIUSERGLYVALLEULEULYGLI 1100  
 DB 3241 GTTACACGACGTGTGCGCATCTGGATGGGCTGTGAGAGCGGGGTCTCTCAAGAG 3300  
 QY 1101 LEUSESERGLYPROLALALEUPROLYSPHEASPLEUHIISERTRPHEANSESERVALILEU 1120  
 DB 3301 CTGAGTGGCCCGGCGCTGCCAAGGACCTGCATAGCACCTTCACTCGGTGTCTGAG 3360  
 QY 1121 PHESETRHAPHEPHEPHEPHEPHEPHEPHEPHEPHEPHEPHEPHEPHEPHEPHEPHE 1140  
 DB 3361 TTCAGACATGACTTCTTCAACAGCAGCAGCAGGAGGCTTGCATTTCACTGTGACCA 3420  
 QY 1141 ALAPHESESERPHEASAPPROGLYILEPROGLINAMGLYSERAARGSERGLYASPSERTP 1160  
 DB 3421 GCAACGTCTCTCATAGACCTGTGGATCCGAGAAATGGAGATCGGAGTGGAGACGTTGG 3480  
 QY 1161 GLUALAGLYASPSERTHVALPHEGLINCYASAPPROGLYTYRALALEUAGLISERLA 1180  
 DB 3481 GAAGCGGCGACTCCACAGTGTTCAGTGTACCTTGTACGCGCTGACAGGAGAGTCA 3540  
 QY 1181 GLUILESERCYVALIYGLIUSANARGPHEPHEPHEPHEPHEPHEPHEPHEPHEPHEPHE 1200  
 DB 3541 GAGATCAGCTGTGTGAAGATCGAAGAACGTTCTTCTGGAGCCAGCCGCGCAACATGC 3600

QY 1201 ILEALAPROCYSGIYGLIYASPLEUETHRGLYPROSERGLYVALILELEUSESERPROASNTYR 1220  
 DB 3601 ATGCTCTCTCGCGGGGAGACCTTACAGACCAATCTGAGTCACTCTTCCACCAATTAC 3660  
 QY 1221 PROGLUPROTYRPROPROGLIYLYGLIUCYASAPTRPILYVALTHRVALLSERPROASPTR 1240  
 DB 3661 CCAGAACCTTACCCGCCAGGAGAGTGTGACTGGAAGTGAACCGTCTCACAGACTAC 3720  
 QY 1241 VALILEALALEUVALPHEANILEPHEANLEULIUPROGLYTYRASPHELEUHIILE 1260  
 DB 3721 GTCATTCGCCCTGTGATTTAACTATTCTTAACCTTGAGCGCTGTGATGACTTCTTCAATATC 3780  
 QY 1261 TYRASPGLIYARGASPSETRLEUSESERPROLEUILEGLYSERPHEPTRYGLYSESGHLEUPRO 1280  
 DB 3781 TACAGCGAGCGGAGCTTCTTCAAGCTCTCTTACAGAAAGCTTCTTATGCTCCAGCTCCA 3840  
 QY 1281 GLVARGILEGLIUSERSESERSEASNSETRLEUVALAPHEARGSERASPALASERVAL 1300  
 DB 3841 GGCGCAATGGAACACAGACAGCAACAGCTTCTTCCGCTTCCGACAGATGATCTGTG 3900  
 QY 1301 SERANALAGLYPHEVALILEASPTYRTHRGJLUSANPROARGJUSERCYSPHEASPRO 1320  
 DB 3901 AGCAATGCTGTGCTTCATTTGACTATACAGAAAACCGCGGAGTCAATGTTTGTATCT 3960  
 QY 1321 GLYSERIILEYASANGLYTHRARGVALGLYSERAPLEUYSLEAGLYSESERVALTHR 1340  
 DB 3961 GGTTCATCAAGAACGGGACAGGAGTGGGTCCACTGAAAGCTGGGCTCTCCGTCAAC 4020  
 QY 1341 TYRTRYCYSHIAGJGLIYTRGLIUNVALGLIUNLYTHSERTRHLEUSESERCYILEUAGLY 1360  
 DB 4021 TACTACGCGACCGGGGCTACGAGATTGAGGGACCTCGAACCTGAGTGCATCTGGGG 4080  
 QY 1361 PROASPGLIYSPROVALITRPAASNAPROARGPROVALCYSTRHAPALPROCYGLIY 1380  
 DB 4081 CCTGATGGAGAACCGCGTGTGGAACAATCCCGGCACTGACAGCCCTCTGGGGGGA 4140  
 QY 1381 GLINTYRVALGLYSERAAPGLYVALILEUSESERPROANTRYPROGLINANTYRTHSER 1400  
 DB 4141 CAGTATGGGTGTGAGAGAGTGTCTGTGCTCCCAACTACCCCAAGATTAACCAAGT 4200  
 QY 1401 GLYGLINILECYSELEUTYRPHENALITHRVALLPROLYSAPTRYVALPHEGLIUNPHE 1420  
 DB 4201 GGAAGATCTGCTTGTATTTTGTTCATCTGTGCCAAGGACATATGTGTGTGGCAAGTTC 4260  
 QY 1421 ALAPHEPHEITHRALALEUASNAPVALIGLIUNVALHISAPGLYHISERGLIHIS 1440  
 DB 4261 GCCTTCTTTCACAGCGGCTTCAAGACGTGTGAGGTTCCAGCGGCCACAGCACAC 4320  
 QY 1441 SERARGLEUSESERSETRLEUSESERGLYSETHIETHRGJLGLIUSERLEUPROLEUVALTHR 1460  
 DB 4321 TCGCGGCTCTCAGCTCTCTCTCGGCTCCCATACAGAGATACAGCTTGTGGCAAC 4380  
 QY 1461 SERANGLINVALILEUILEYSPHESEVALYSEGLYLEUVALPROLALARGGLYPHEHIS 1480  
 DB 4381 TCCATATGATGTTCTCATTAAGTTCAGGCGCAAAAGCTCTGCAGCAGCAGAGGCTTCCAC 4440  
 QY 1481 PHEVALTYRGLIUNVALPROARGTRHSEVALATHRGJUSERSESERVALPROGLUPRO 1500  
 DB 4441 TTTGTCTACCAAGGAGTTCCTCGAACAAGCGCCACGAGTGCAGCTGTGTGCCGAACCC 4500  
 QY 1501 ARGTRYGLIYASAGLEULYSERAASPHESEVALIGLIUNVALILEVALARGPHEGLIUCYS 1520  
 DB 4501 CGTATGACAGAGAGCTGGGAGTGTCTCGATGGAGGACATGCTTGTGAAATGC 4560  
 QY 1521 ASNSERGLYTYRALALEUAGLISERPROGLIUNILEGLUCYSELEUPROVALPROGLIYALA 1540  
 DB 4561 AACTCCGCGATATGCTCGAGGAGTCCGCAAGATTCAGTGTCTCTGTGTGCTGGGGCC 4620  
 QY 1541 LEUVALGLINTRPAENVALSERALAPROTHRCYVALIVALPROCYSGIYGLIYASNTLEUTHR 1560  
 DB 4621 TTGGCCCAATGAAATGTCTCAGCGCCACAGTGTGTGTGCTCGTGTGAGGACCACTCACA 4680  
 QY 1561 GLIUNARGGLYTHRIILEUSESERPROGLYPHEPROGLUPROTYRLEUASNSETRLEUASN 1580

```
Db 4681 GAGCGAGGGGAGCCATCTCTGTCCTCCCTGGCTTCCAGAGCCCTACTCAACGCTCTCAAC 4740
Qy 1581 CyValATTrpValIleValValProGluGlyAlaGlyIleGlnIleGlnValValSerPhe 1600
Db 4741 TGGTGTGGAAATCGTGTGTCCCGAAGCGCTGGCATCAGATCCAAAGTTGTGAGTTT 4800
Qy 1601 ValThrGluGlnAsnTrpAspSerLeuGluValPheAspGlyAlaAspAsnThrValThr 1620
Db 4801 GGGACAGAGCACAACCTGGAGCTCGCTGGAGATTTGATGGGAGATTAACACTGTAACT 4860
Qy 1621 MetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerAsnGln 1640
Db 4861 ATGCTGGGGAGTTTCTCAGAAACAACCGTGGCTGCTTCCGAACAGACTCCAAACAG 4920
Qy 1641 LeuThrLeuHisPheThrSerAspIleSerValSerAlaAlaGlyPheHisLeuGluThr 1660
Db 4921 CTCTACCTTCACTTCTTACTCAGATATACAGCGATCTGACAGCTGGCTTCCACTTGGAGTAC 4980
Qy 1661 LysThrValGlyLeuSerSerCybProGluProAlaValProSerAsnGlyValIleThr 1680
Db 4981 AAAACGTGGGGCTGAGCAGATTGTCCGAACTGTGTGCCAGTTAACGGGGTGAAGACT 5040
Qy 1681 GlyGluArgThrLeuValAsnAspValValSerPheGlnCybGluProGlyThrValLeu 1700
Db 5041 GGGAGCGCTACTGTGTGAATGATGTGTGTCTTTCAGTGTGAGCCGGGATATGCCCTC 5100
Qy 1701 GlnGlyHisAlaHisIleSerCybMetProGlyThrValArgArgTrpAsnThrProPro 1720
Db 5101 CAGGGCCACGCCCAATCTCTGCATGCCGGAACAATGGGCGATGAACCTACCTCTC 5160
Qy 1721 ProLeuCybIleAlaGlnCybGlyGlyThrValGluGluMetGluGlyValIleLeuSer 1740
Db 5161 CCACCTGTATATGACAGTGTGGGGGAAACAGTGGAGAGATGGAGGGGGTGTACTCTGAGC 5220
Qy 1741 ProGlyPheProGlyAsnThrProSerAsnMetAspCybSerTrpValIleAlaLeuPro 1760
Db 5221 CCGGGCTTCCAGGCAACTACCCCAATACATGAGACTGTCTCGGAAATATGACACTGCC 5280
Qy 1761 ValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrGluProAsnHisAspThr 1780
Db 5281 GTGGGCTTTGGAGCTCACTACATCCAGTTCTGAACTTCCACGAGCCCAACACGACTAC 5340
Qy 1781 IleGluIleArgAsnGlyProGlyThrGluThrSerArgMetLeuValArgPheSerGlySer 1800
Db 5341 ATGAAATCCGGAATGGCCCCCTATGAGACCAAGCCGATGATGGGAAGATTCTAGTGAAAC 5400
Qy 1801 GluLeuProSerSerLeuLeuSerThrSerHisGlyThrThrValTyrPheHisSerAsp 1820
Db 5401 GAGCTTCAAGCTCCCTCCCTCCACGTCCTCCACGAGACCACTGTATTTCCACAGCCGAC 5460
Qy 1821 HisSerGlnAsnArgProGlyPheIleValLeuGluThrGlnAlaTyrGluLeuGlnGly 1840
Db 5461 CACTCCCAAGATGGGCGAGATTCAAGCTGAGATATCAGGCCATATGAATCTTCAAGAGTGC 5520
Qy 1841 ProAspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTyrTrpAsnValGlyGln 1860
Db 5521 CAGACCCAGAGAGCTTTGGCAATGGCATTTGAGAGGAGAGCTGGCTACAACTGGAGACA 5580
Qy 1861 SerValThrPheGluCybLeuProGlyTyrGlnLeuThrGlyHisProValLeuThrCys 1880
Db 5581 TCAGTACCTTCCAGAGCTCCCGGGGATCAATTGACTGGCCACCTGCTCTCCAGTGT 5640
Qy 1881 GlnHisGlyThrAsnArgAsnTrpAspHisProLeuProValCysGluValProCysGly 1900
Db 5641 CAACATGGGACCAACCGGAATGGGAGCAACCCCTGGCCCAAGTGGAGTCCCTTGTGGC 5700
Qy 1901 GlyAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProThrSer 1920
Db 5701 GGGACATCACTTCTTCCAAAGCACTGTACTCCCGGGGTCTCCAGCCCTGACTCC 5760
Qy 1921 SerSerGlnAspCybValThrLeuIleThrValProIleGlyHisGlyValArgLeuAsn 1940
Db 5761 AGCTCCAGAGACTGTGTCTGCTGATCACGTCGCCCATTTGGCATGGCGTCCGCTCAAC 5820
Qy 1941 LeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTrpAspGlyProGln 1960
Db 5821 CTAGCCTGTGACAGACAGAGCCCTCTGGAGATTTCAATACATCAGATCGGAGTGGGACAG 5880
Qy 1961 GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaIleValThrValGlnSer 1980
Db 5881 CAACAGACCAACAGGCTCGCGCTTCAACCGGAGATGGCCAAAGAAACATGTCAGACT 5940
Qy 1981 SerSerAsnGlyValLeuLeuLysPheHisAlaGlyAlaIleThrGlyGlyIlePheAla 2000
Db 5941 TCATCCAAACAGATCTCTGTCAAGTTCCACCGTATGAGCCACAGCGGGGGATCTTCCGC 6000
Qy 2001 IleAlaPheSerAlaTyrProLeuThrLysCybProProProThrIleLeuProAsnAla 2020
Db 6001 ATAGCTTCTCGCTTATCACTCAACAAATCCCTCTCCACATCTCCCAAGCC 6060
Qy 2021 GluValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrArgCysLeu 2040
Db 6061 GAAGTCGTCAAGAAATGAAGAAATTAATATAGGTGACATGTACGCTACAGATGCCCTC 6120
Qy 2041 ProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuGlyThrThrLeuGln 2060
Db 6121 CTTGGCTTACTTACTGTGGGAATGAATTTCTGACTGTCAAACTTGAACTTACCTGACG 6180
Qy 2061 PheGluGlyProProProIleCysGlyValHisCybProThrAsnGluLeuLeuThrAsp 2080
Db 6181 TTTGAAGACCAACCCCGATATGTAAAGTACGCTGCAACAAATGAGCTTCTACAGAC 6240
Qy 2081 SerThrGlyValIleLeuSerGlnSerThrProGlySerThrProGlnPheGlnThrCys 2100
Db 6241 TCCAGAGCGTGATCTGTAGCCAGAGCTACCTGAGAGCTATCCCGAGTTCCACACTGC 6300
Qy 2101 SerTrpLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluThrPheLeu 2120
Db 6301 TCTTGGCTGTGAGAGTGAGACCCGACTATTAACATCTCCCTCACAGTGGAGTACTTCTC 6360
Qy 2121 SerGluLysGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeu 2140
Db 6361 AGCAGAGCAATATGATGATTTGAGATTTGATGTGCTCATCAGACAGAGACTCTG 6420
Qy 2141 LeuLysAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerAspAsnSer 2160
Db 6421 CTGAAGCCCTCAATGGGAATTACTACGCTCCCTGATTTCTACACACTCAAGCACTCT 6480
Qy 2161 ValTyrLeuArgTrpSerSerAspHisAlaTyrAsnArgLysGlyPheLysIleArgTyr 2180
Db 6481 GTGTACCTGGTGTGTGATCTGATCAAGCTTCAATGGAAGGCTTCAAGATCCGCTAT 6540
Qy 2181 SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr 2200
Db 6541 TCAGCCCTTACTGACCTCTGCGAGGCTCACTCCATGAGCTTCACTTCAAGCCAGAC 6600
Qy 2201 SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly 2220
Db 6601 AGCACCCAGCCCGGGGCTCATCTGCTGTCGACAGCGGCTACACCCCTGTGGGA 6660
Qy 2221 HisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuThrSerGluAlaIle 2240
Db 6661 CACAGCTGGCAATGTGTACCGGCAACCCCAAGGCTTACCACTGTGGAGCGCAAGCATTC 6720
Qy 2241 ProLeuCybGlnAlaLeuSerCysGlyLeuProGluAlaProLysAsnGlyMetValPhe 2260
Db 6721 CTTCTGTGACACTTTTCTGTGTGGCTTCTGAGGCCCCCAAGATGAATGTGTGTT 6780
Qy 2261 GlyLysGluThrThrValGlyThrLysAlaValTyrSerCysSerGluGlyTyrHisLeu 2280
Db 6781 GGCAGAGAGTACACATGGGAAACAAAGCCGTGTACAGCTGACATGGAAGCTTACCACTC 6840
Qy 2281 GlnAlaGlyValAlaGluAlaThrAlaGluCysLeuAspThrGlyLeuThrSerAsnArgAsn 2300
Db 6841 CAGGAGGGGCTTGGGCCCTGCAAGTGTCTGAGACAGGCGCTATGAGCAACCGCAAT 6900
```

|    |      |   |      |    |      |   |      |
|----|------|---|------|----|------|---|------|
| Qy | 2301 | ValProProGlnCyValProValThrCysProAspValSerSerIleSerValGlnHis       | 2320 | Qy | 2661 | SerLeuProHisCysSerGlyThrSerValGlyValCysGlyAspProGlyIleProAla    | 2680 |
| Db | 6901 | GTCCACACACAGTGTGTCCCTGTACTGTCTGTATGTCACTAGCATCAGGTGGAGAT          | 6960 | Db | 7981 | TCCCTCCCTCAGCTCAGAGACAGCGTGGAGTTGGCGTGAACCTGGGATCCCGCT          | 8040 |
| Qy | 2321 | GlyArgTrpArgLeuIlePheGlnThrGlnTyrglnPheGlnIleGlnLeuMetLeuIle      | 2340 | Qy | 2681 | HisGlyIleArgLeuGlyAspSerPheAspProGlyThrValIleArgPheSerCysGlu    | 2700 |
| Db | 6961 | GGCCGATGGAGGCTTATCTTGAGACACAGTATCAGTTCAGAGCCCACTGATGTCTATC        | 7020 | Db | 8041 | CATGGCATCCGTTTGGGGGACAGCTTGAATCCAGGCACTGTGATGCCCTTCAGCTGGAA     | 8100 |
| Qy | 2341 | CysAspProGlyYrTrpTrpTrpThrGlyGlnArgValIleArgCysGlnIleAsnGlyLys    | 2360 | Qy | 2701 | AlaGlyIleValLeuArgGlySerSerGlnAspThrCysGlnIleAsnGlySerTrpSer    | 2720 |
| Db | 7021 | TGTGACCTGGCTACTACTATACGTGGCAAGGGTCATCCGTGCAGGCCCAATGGCAAA         | 7080 | Db | 8101 | GCTGGCAAGTGTCCGGGAGTGTCTGACAGCGACCTGTCAAGCCATAGCTGTGGAGC        | 8160 |
| Qy | 2361 | TrpSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGlnLeuProIle      | 2380 | Qy | 2721 | GlySerGlnProGlnCysGlyValIleSerCysGlyAspProGlyThrProSerAsnAla    | 2740 |
| Db | 7081 | TGGAGCTTCGGGAGCTTACGCCCACTGCCGAATCATCTCTGTGGAGAGCTCCCAAT          | 7140 | Db | 8161 | GGCTCGAGGCTTAGTGATGAGTATCTTTGGGGAACCTGGGACTCCAGATGAATGCC        | 8220 |
| Qy | 2381 | ProProAsnGlyHisArgIleGlyThrLeuSerValTyrglyAlaThrAlaIlePheSer      | 2400 | Qy | 2741 | ArgValAlaPheSerAspGlyLeuValPheSerSerIleValTyrglyCysArgGlu       | 2760 |
| Db | 7141 | CCCCCATGGCCACCGCATCGAACACTGTCTGTCTAGGGGCAACAGCCATCTTCTCC          | 7200 | Db | 8221 | CGAGTTGTGATGATGAGCTGTGTTTCTCCAGCTCTATCGTATGAGTCCGGGAA           | 8280 |
| Qy | 2401 | CysAsnSerGlyYrTrpThrLeuValGlySerArgValIleArgGluCysMetAlaAsnGlyLeu | 2420 | Qy | 2761 | GlyTrpTrpAlaThrGlyLeuLeuSerArgHisCysSerValAsnGlyTrpTrpThrGly    | 2780 |
| Db | 7201 | TGCAATTCGGGATACACACTGTGTGGCTCCAGGGTGCCTGAGTGCATGGCCAAATGGGCTC     | 7260 | Db | 8281 | GGATACTACGCACAGGCTGTCTAGCGTACTGTCTGGTCAATGGTACTGGACAGGC         | 8340 |
| Qy | 2421 | TrpSerGlySerGlnValArgCysLeuAlaGlyHisCysGlyYrTrpProGluProIleVal    | 2440 | Qy | 2781 | SerAspProGlnCysLeuValIleAsnGlyAspProGlyIleProAlaAsnGlyLeu       | 2800 |
| Db | 7261 | TGGAGTGTCTGAGTACCTCCGCTGCTTGTGGACACTGTGGAGCTCCTGAGCCCATGTCTC      | 7320 | Db | 8341 | AGTACCTCAGTACCTCCGTCTCATTAACCTGTGTGATCCCTGGAGTCCAGCCAAATGGCTT   | 8400 |
| Qy | 2441 | AsnGlyHisIleAsnGlyGlnAsnTyserTyArgGlySerValValTyrglnCysAsn        | 2460 | Qy | 2801 | ArgLeuGlyAsnAspPheArgTyArgIleThrValThrTyrglnCysValProGlyYr      | 2820 |
| Db | 7321 | AACGGACATCAATGGGGAGAACACAGCTACCGGGGCAGTGTGTATCAATGCAT             | 7380 | Db | 8401 | CGGCTGGGCAATGACTCAGGTACAAACAATGTGACATATGATGTGCTCCGTGCTAT        | 8460 |
| Qy | 2461 | AlaGlyPheArgLeuIleGlyMetSerValArgIleCysGlnIleAspHisIleTrpSer      | 2480 | Qy | 2821 | MetMetGlnSerHisArgValSerValLeuSerCysThrTrpAspArgTrpTrpAsnGly    | 2840 |
| Db | 7381 | GCTGCTTCCGCTGATCGGATGTCTGTGGCATCTGCCAGAGATCATCATCAGTGTCTG         | 7440 | Db | 8461 | ATGATGAGTACATTAAGATATCTGTGTAGCTGACCAAGGACCGGAAATGGAATGGA        | 8520 |
| Qy | 2481 | GlyLysThrProPheCysValProIleThrCysGlyHisAspProGlyAspProValAsnGly   | 2500 | Qy | 2841 | ThrLysProValCysIleAlaLeuMetCysLysProProLeuIleProAsnGlyLys       | 2860 |
| Db | 7441 | GGCAAGACCCCTTCTGTGTGCCAATTAACCTGTGGACACCAAGGCAACCTGTCAACGGC       | 7500 | Db | 8521 | ACCAAGCCGCTTCGCAAGCTCATATGTGCAAGCACTCCGCTCAATCCCAATGGGAAG       | 8580 |
| Qy | 2501 | LeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValLysPheValCysAsnProGly      | 2520 | Qy | 2861 | ValValGlySerAspPheMetTrpGlySerSerValThrTyrlaCysLeuGlnGlyYr      | 2880 |
| Db | 7501 | CTCATCTAGGGGTAACAGTTTAACCTCAACAGATGTGCAAGTTGTTTGCACCCCTGG         | 7560 | Db | 8581 | GTGGTGGGATCTGACTCATGTGGGCTCAAGTGTGACTTATGCTCTCGAGGGGTAC         | 8640 |
| Qy | 2521 | TyrMetAlaGlnGlyAlaIleArgSerGlnCysLeuAlaSerGlyGlnTrpSerAspMet      | 2540 | Qy | 2881 | GlnLeuSerLeuProAlaValPheThrCysGlnGlyAsnGlySerTrpThrGlnGlnLeu    | 2900 |
| Db | 7561 | TATATGGCTGAGGGGCTGTAGTCCCAATGCCCTGGCCAGCGGCAATGAGTGAATG           | 7620 | Db | 8641 | CAGCTCTCCCTGCCCCGGGTGTCACTGTGAGGAATGAGTCTTGAGCCGAGAGCTG         | 8700 |
| Qy | 2541 | LeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGlnAsnSerValArg      | 2560 | Qy | 2901 | ProGlnCysPheProValPheCysGlyAspProGlyValProSerArgGlyArgArgGlu    | 2920 |
| Db | 7621 | CTGCCACCTGCAAGATATCACTGATCAGATCTCTGGACACCAAGAAATATGATGTCT         | 7680 | Db | 8701 | CTTCAGTGTTCCTGTGTCTGTGGGATCTGTGTCTCCGTCCGTGGGAGAGAGAG           | 8760 |
| Qy | 2561 | GlnValHisAlaSerGlyProHisArgPheSerPheGlyYrThrThrValSerTyArgCys     | 2580 | Qy | 2921 | AspArgGlyPheSerTyArgSerSerValSerPheSerCysHisProProLeuValLeu     | 2940 |
| Db | 7681 | CAGGTCCACGCGAGGGGCCCGACAGTTCAGCTTCGGCACCACTGTGTCTTACCGGTC         | 7740 | Db | 8761 | GACGAGAGCTTCTCTACAGTATCTGTCTCTCCGTTCCTGCAATCCCTCTGTGTCTG        | 8820 |
| Qy | 2581 | AsnHisGlyPheTyTrpLeuGlnGlyThrProValLeuSerCysGlnGlyAspGlyYrTrp     | 2600 | Qy | 2941 | ValGlySerProArgArgPheCysGlnSerAspGlyYrThrTrpSerGlyYrGlnProSer   | 2960 |
| Db | 7741 | AACCAAGGCTTCTACTCTCTGGGCAACCCAGTGTCACTGCAAGGAGATGGACATGG          | 7800 | Db | 8821 | GTGGGCTCTCCACCAAGTGTGTGGCACTGAGTGGACATGAGGTGCAACCGCCAGC         | 8880 |
| Qy | 2601 | AspArgProArgProGlnCysLeuLeuValSerCysGlyHisAspProGlyYrTrp          | 2620 | Qy | 2961 | CysIleAspProThrIleLeuThrThrCysAlaAspProGlyValProGlnPheGlyIleGln | 2980 |
| Db | 7801 | GACCTTCCCGCCCGCCAGTGTCTTGTGTGTCTGTGTGCAATCCGGGTCTCCCGCTTAC        | 7860 | Db | 8881 | TGCATTAATCCGACCTTGAACCACTGTGGGACCTGTGTGTGCAACAGTTTGGATACAG      | 8940 |
| Qy | 2621 | SerGlnMetSerGlyAspSerTyTrpValGlyAlaValAlaArgTySerCysIleGly        | 2640 | Qy | 2981 | AsnAsnSerGlnGlyTyrglnValGlySerThrValLeuPheArgCysGlnGlyYr        | 3000 |
| Db | 7861 | TCCCAAGATGTCTGGAGACAGTATACTGTGGAGACAGTGTGCGTTACAGCTGCATCGGC       | 7920 | Db | 8941 | AACATTCCTACGAGGCTACAGGTTGAGAACACAGTCTCTTCGGTTGTCAAAAAGCTAC      | 9000 |
| Qy | 2641 | LysArgThrLeuValGlyAsnSerThrArgMetCysGlyLeuAspGlyHisIleTrpThrGly   | 2660 | Qy | 3001 | LeuLeuGlnGlySerThrThrArgThrCysLeuProAsnLeuThrTrpSerGlyYrPro     | 3020 |
| Db | 7921 | AAGCTACTCTGTGGGAAACAGACCCGCACTGTGTGGGCTGTGATGACACTGAGCTGGC        | 7980 | Db | 9001 | CTGCTTACGAGGCTCCACACAGGACCTGTCTCCAAACCTGTACCTGGAGTGAACCCCA      | 9060 |
| Qy | 2661 | ProAspCysValProHisHisCysArgGlnProGlnThrProThrHisAlaAsnValGly      | 3040 | Qy | 3021 | ProAspCysValProHisHisCysArgGlnProGlnThrProThrHisAlaAsnValGly    | 3040 |

DB 9061 CCTGACGTGTCCCCCACTGCAGCAGCAGCAACGATGCACGTCGGG 9120  
QY 3041 AAlenuspLeupProserMetGlyThrThreuleTherProAlaagaAlaSerPro 3060  
DB 9121 GCCCTGATTTGCTCTCCATCAAGGGCTACACGCTCATTTCTCTGCAGAGGGCTTCTCC 9180  
QY 3061 SerArgValAlaProSerThralaProAlaArgMetAlaAlaGlyGlnAlaSerArg 3080  
DB 9181 TCAAGGGTGGCTCGACACCGACCTCGAAGCGGATGGCAGCTGACAGGCAAGCCGC 9240  
QY 3081 ProSerAlaTrpArgSerGlyProValGlyAspProSerThreuproGlySerHISArg 3100  
DB 9241 CCATCTGCTCGAGAGTCCGGCCCACTGGGAGAGCCCATCACTGCCCCGGAGCCACCGC 9300  
QY 3101 SerProlySerPro 3104  
DB 9301 TCACCCAGAGCT 9312

RESULT 3  
ID ADH71137 standard; DNA; 10136 BP.  
XX ADH71137;  
AC ADH71137;  
XX 25-MAR-2004 (first entry)  
DT 25-MAR-2004 (first entry)  
XX  
DE Human gene of the invention NOVab SEQ ID NO:33.  
XX  
KM de; gene; human; cytostatic; immunomodulator; neuroprotective; nootropic;  
KM anorectic; antidiabetic; antimicrobial; antihypertensive; gene therapy;  
KM vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;  
KM obesity; diabetes; infectious disease; metabolic syndrome X;  
KM dyslipidemia.  
XX  
OS Homo sapiens.  
XX  
PN WO2003102155-A2.  
XX  
PD 11-DEC-2003.  
XX  
PF 03-JUN-2003; 2003WO-US017430.  
XX  
XX 03-JUN-2002; 2002US-0385120P.  
PR 04-JUN-2002; 2002US-0385784P.  
PR 05-JUN-2002; 2002US-0386041P.  
PR 05-JUN-2002; 2002US-0386047P.  
PR 06-JUN-2002; 2002US-0386376P.  
PR 06-JUN-2002; 2002US-0386453P.  
PR 06-JUN-2002; 2002US-0386864P.  
PR 07-JUN-2002; 2002US-0387016P.  
PR 07-JUN-2002; 2002US-0386796P.  
PR 07-JUN-2002; 2002US-0386816P.  
PR 07-JUN-2002; 2002US-0386931P.  
PR 07-JUN-2002; 2002US-0386942P.  
PR 07-JUN-2002; 2002US-0386971P.  
PR 07-JUN-2002; 2002US-0387262P.  
PR 08-JUN-2002; 2002US-0296960P.  
PR 10-JUN-2002; 2002US-0387400P.  
PR 10-JUN-2002; 2002US-0387535P.  
PR 11-JUN-2002; 2002US-0387610P.  
PR 11-JUN-2002; 2002US-0387625P.  
PR 11-JUN-2002; 2002US-0387634P.  
PR 11-JUN-2002; 2002US-0387668P.  
PR 11-JUN-2002; 2002US-0387686P.  
PR 11-JUN-2002; 2002US-0387702P.  
PR 11-JUN-2002; 2002US-0387836P.  
PR 11-JUN-2002; 2002US-0387859P.  
PR 12-JUN-2002; 2002US-0387933P.  
PR 12-JUN-2002; 2002US-0387934P.  
PR 12-JUN-2002; 2002US-0387960P.  
PR 12-JUN-2002; 2002US-0388022P.  
PR 12-JUN-2002; 2002US-0388096P.

PR 13-JUN-2002; 2002US-0389123P.  
PR 14-JUN-2002; 2002US-0389118P.  
PR 14-JUN-2002; 2002US-0389120P.  
PR 14-JUN-2002; 2002US-0389144P.  
PR 14-JUN-2002; 2002US-0389146P.  
PR 14-JUN-2002; 2002US-0389146P.  
PR 17-JUN-2002; 2002US-0389729P.  
PR 17-JUN-2002; 2002US-0389742P.  
PR 18-JUN-2002; 2002US-0389844P.  
PR 19-JUN-2002; 2002US-0390066P.  
PR 19-JUN-2002; 2002US-0390209P.  
PR 21-JUN-2002; 2002US-0390763P.  
PR 17-JUL-2002; 2002US-0396706P.  
PR 06-AUG-2002; 2002US-0401628P.  
PR 09-AUG-2002; 2002US-0402156P.  
PR 09-AUG-2002; 2002US-0402566P.  
PR 12-AUG-2002; 2002US-0402389P.  
PR 12-AUG-2002; 2002US-0402786P.  
PR 12-AUG-2002; 2002US-0402816P.  
PR 12-AUG-2002; 2002US-0402821P.  
PR 12-AUG-2002; 2002US-0402832P.  
PR 13-AUG-2002; 2002US-0403448P.  
PR 13-AUG-2002; 2002US-0403459P.  
PR 13-AUG-2002; 2002US-0403531P.  
PR 13-AUG-2002; 2002US-0403532P.  
PR 13-AUG-2002; 2002US-0403563P.  
PR 13-AUG-2002; 2002US-0406317P.  
PR 15-AUG-2002; 2002US-0406317P.  
PR 26-AUG-2002; 2002US-0406182P.  
PR 26-AUG-2002; 2002US-0406355P.  
PR 27-AUG-2002; 2002US-0406240P.  
PR 12-SEP-2002; 2002US-0410084P.  
PR 20-SEP-2002; 2002US-0412528P.  
PR 23-SEP-2002; 2002US-0412731P.  
PR 30-SEP-2002; 2002US-0414801P.  
PR 30-SEP-2002; 2002US-0414839P.  
PR 30-SEP-2002; 2002US-0414840P.  
PR 30-SEP-2002; 2002US-0414954P.  
PR 09-OCT-2002; 2002US-0417186P.  
PR 09-OCT-2002; 2002US-0417406P.  
PR 23-OCT-2002; 2002US-0420639P.  
PR 28-OCT-2002; 2002US-0421156P.  
PR 31-OCT-2002; 2002US-0422690P.  
PR 01-NOV-2002; 2002US-0423130P.  
PR 05-NOV-2002; 2002US-00423798.  
PR 05-NOV-2002; 2002US-0423798P.  
PR 12-NOV-2002; 2002US-0425453P.

(CURA-) CURAGEN CORP.  
XX  
PA  
XX  
PI Alsbrook JP, Alvarez E, Anderson DW, Boldog FL, Casman SJ;  
PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;  
PI Eitenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;  
PI Gusev VI, Herrmann UL, Li W, Kekuda R, Li U, Liu X, Macdougall JR;  
PI MacIsachlan T, Malysankar UM, Mezick AJ, Millet I, Mishra VS;  
PI Padigaru M, Patcurajan M, Pena CE, Peyman JA, Raha D, Raselli L;  
PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shimkets RA;  
PI Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;  
PI Zhong H;  
XX  
XX WPI; 2004-081935/08.  
DR P-PSDB; ADH71138.  
XX  
XX  
PT New NOVX polypeptides and nucleic acid molecules useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or  
PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.  
XX  
XX  
PS Example 4; SEQ ID NO 33; 1880bp; English.  
XX  
XX The invention relates to a novel isolated polypeptide (NOVX). A  
CC polypeptide of the invention has cytostatic, immunomodulator,  
CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and  
CC antihypertensive activity, and may have a use in gene therapy, and as a  
CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising



```
QY 621 LeuGlnSerAspHisCysLeuAspProGlyYlleProValAsnGlyGlnArgHisGlyAsn 640
DB 1861 CTGCACATCAGACCACTGCTGATCCAGAAATCCAGTAATAGGACACGGTCATAGGAAAT 1920
QY 641 AspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrThrLeuSerAsp 660
DB 1921 GACTTCTACGTGGGGCGCTGGTGAACCTTCAGCTGTGACTGGGCTTACATTAAGTAC 1980
QY 661 GlyGluProLeuGluCysGluProAsnPheGlnTTPSerArgAlaLeuProSerCysGlu 680
DB 1981 GGGAGACCTCTGAGTGTGAGCCCAACTTCAGTGGAGCCGGGCCCTGCCAGTGTGGAA 2040
QY 681 AlaLeuCysGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSerProGlyPhePro 700
DB 2041 GCTCTGTGGTGGCTTCATTCAGAGCTCCAGTGGACCATCTTGCCAGGGTTCCTCT 2100
QY 701 AspPheTyrProAsnAsnLeuAsnCysThrTrrIleIleGlnThrSerHisGlyLysGly 720
DB 2101 GACTTCTACCCCAACACCTTGAACCTGACCTGGATTTATCGAAACATCTCATGGCAAGGT 2160
QY 721 ValPhePheThrPheIleThrPheHisLeuGluSerGlyHisAspTyrIleLeuIleThr 740
DB 2161 GTGTTCTTCACTTTCACACCTTCACCTGGAAAGTGGCCATGACTCTCTCATCAT 2220
QY 741 GluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760
DB 2221 GAGAACGGCAGCTTACCCAGCCCTGAGGAGCATGAGTCTGGATCTGGCCGACGCTCCC 2280
QY 761 IleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780
DB 2281 ATCAGCGCTGGGCTCTTATGGCACTTCACTGCCCCAGGTCGCTCATCTCTGATTTCTCC 2340
QY 781 MetSerTyrGluGlyPheAsnIleThrPheSerGluTyrAspLeuGluProCysGluGln 800
DB 2341 ATGTCATATGAAAGGATTCACATCACTTCTCAGAGAACACATTTGAGCCCTTGAGAG 2400
QY 801 ProGluValProAlaTyrSerIleArgLysGlyLeuGlnPheGlyValGlyAspThrLeu 820
DB 2401 CCCAGAGTCCCGAGCTTACAGACATCCGGAAAGGGCTTGAGTTGGCTGGGGCGACACTTG 2460
QY 821 ThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIleThrCysLeuGly 840
DB 2461 ACCCTTCTCGTCTCCCGGGATCCGCTGAGAGGACCGCCGCAATCAAGTCCCTGGGG 2520
QY 841 GlyArgArgArgLeuTTPSerSerProLeuProArgCysValAlaGluCysGlyAsnSer 860
DB 2521 GGGAGAGGGCGCTGTGGAGCTCGCTGCAAGGTGTGTGCTGAGTGGGAATTCA 2580
QY 861 ValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTyrAsnAsnAsn 880
DB 2581 GTCAACAGGCACTCAGGGATCTTGTGTGCTCCCAACTTTCGTGTGAACATTAACAT 2640
QY 881 HisGluCysIleTyrSerIleGlnThrGlnProGlyLysGlyIleGlnLeuLysAlaArg 900
DB 2641 CATGAATGATCTACTACTCATCCAGACCCAGCGAAGGAATTCAGTGAACCCAGG 2700
QY 901 AlaPheGluLeuSerGluGlyAspValLeuLysValTyrAspGlyAsnAsnAsnSerAla 920
DB 2701 GCAATTGAACCTCCCAAGAGAAATGTCTCAAGGTTTATGATGGCAACAACAATCCGCC 2760
QY 921 ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAsnSerThrSer 940
DB 2761 CGTTGCTGGAGTTTTCGCCATCTTGAGATGAGGGGATGACTTTGAACAGCAATCC 2820
QY 941 SerSerLeuTTPLeuAspPheIleThrAspAlaGluAsnThrSerLysGlyPheGluLeu 960
DB 2821 AACAGCTGTGGCTGATTCATCACTGATGCTGAACCAACAGCAAGAGGCTTTGAATG 2880
QY 961 HisPheSerSerPheGluLeuLeuLysCysGlyAspProGlyThrProLysPheGlyTyr 980
DB 2881 CACTTTTCAAGCTTTGAACCTCAATAATGTAGAGCCAGAGAACCCCAAGTTTGGCTAC 2940

QY 981 LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCysAspProGly 1000
DB 2941 AAGGTCATGATGAATGAATGATTTTGGAGGAGACTCCGTCTCTTCAAGCTGTACCTGGA 3000
QY 1001 TyrSerLeuArgGlySerGluLeuLeuCysLeuSerGlyGluArgArgThrTrpAsp 1020
DB 3001 TACAGCTCTCGGGGATGATGAGAGCTGTGTCTGATGGAGAGAGCGCCGAGCTGGAG 3060
QY 1021 ArgProLeuProThrCysValAlaGluCysGlyGlyThrValArgGlyGluValSerGly 1040
DB 3061 CGGCTCTGCCACCTGTCTCCCGAGATGTGAGGAGACATGAGAGAGAGGTGTGGGG 3120
QY 1041 GlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIleThrThr 1060
DB 3121 CAGGTGCTGTCAACCCGGGATTCAGCTCTTATGAACAACATCTCACTGATCTGGACC 3180
QY 1061 IleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGluGlu 1080
DB 3181 ATCGAAGCAGAGCGCGCTGACACATGGAGCTACACTTCTGTGTGTGGACACAGAGAG 3240
QY 1081 ValHisAspValLeuArgIleThrAspGlyProValGluSerGlyValLeuLeuLysGlu 1100
DB 3241 GTTCACAGAGTCTGCGCATCGGATGGGCTGTGGAGAGCGGGTTCCTGTAAGAG 3300
QY 1101 LeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValIleGln 1120
DB 3301 CTGAGTGGCCGGCTCTCCCAAGACCTGATGACCTTCAACTGCACTCGGTGCTCGAG 3360
QY 1121 PheSerThrAspPhePheThrSerLysGlnGlyPheAlaIleGlnPheSerValSerThr 1140
DB 3361 TTCAGCACTGACTTCTTACACAGACAGAGGCTTTGCCATTCAATTTTCAGTGTCCA 3420
QY 1141 AlaThrSerCysAsnAspProGlyYlleProGlnAsnGlySerArgSerGlyAspSerTrp 1160
DB 3421 GCACGTCTCTCAATGACCTCGGATCCCGCAAGATGGAGTGGATGGATGGACAGTTGG 3480
QY 1161 GluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeuGlnGlySerAla 1180
DB 3481 GAACCGCGGCACTCAACAGTGTTCACAGTGTACCTTGCTACCGGCTGACGGAGAGTGA 3540
QY 1181 GluIleSerCysValLysIleGluAsnArgPhePheTrpGlnProSerProProThrCys 1200
DB 3541 GAGATCAGCTGTGTGAAGATCGAAGAACAGGTTCTTGGCAGCCGACCAACTGC 3600
QY 1201 IleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr 1220
DB 3601 ATCCCTCCCTGGGGGAGACTGACAGACCATCTGGAGTCACTCTTCAACCAAAATTAC 3660
QY 1221 ProGluProTyrProProGlyLysGluCysAspTrpLysValThrValSerProAspTyr 1240
DB 3661 CCAGAACCTTACCCGCGCAGCAAGAGTGTGACTGAAAGTGCCTGTTCACCAAGACTAC 3720
QY 1241 ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle 1260
DB 3721 GTCAATGCCCTGGATTTTAACATCTTAACTTGAACCTGGACTGTATGACTTCTCCATATC 3780
QY 1261 TyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerGlnLeuPro 1280
DB 3781 TACGAGGAGGAGGACTCTTCAACCTCTCATAGGAAGCTTCTATAGGCTTCCAGCTCCA 3840
QY 1281 GlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal 1300
DB 3841 GGGCGGATTTGAAGCAGACCAAGCCTTCTCTGCTTCCGAGCGATCATCTGTG 3900
QY 1301 SerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCysPheAspPro 1320
DB 3901 AGCAATGCTGGCTTGTCTATTAATACTAACAAGAAACCCCGGAGATGATGTTTATCTCT 3960
QY 1321 GlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeuGlySerSerAlaThr 1340
DB 3961 GGTTTCATCAAGACGGCACACGGGTGGGTCCGACTGAAGCTGGCTCTCTCCCTCAAC 4020
QY 1341 TyrTyrCysHisGlyGlyTyrGluValGluGlyThrSerThrLeuSerCysIleLeuGly 1360
```

```
Db 4021 TACTACTGCCACGGGGGCTACGAGATTGAGGGGACCTGACCCCTGAGCTGATCCTGGG 4080
Qy 1361 ProaapgllylvprovalITPaaanaProaigrprovalCyethralaProCyagly 1380
Db 4081 CCTGATGGGAAGCCCGTGTGAAACAATCCCGGCCAGTCTGCACAGCCCTGTGGGGGA 4140
Qy 1381 GlnTyValaGlySeraspGlyValaVlaLeuSerProaenTyProGlnaasnTyThSer 1400
Db 4141 CAGnTnTGSGGTTCCGACGAGTGGTCTTGTCTCCCACTACCCCAAGACTACCACT 4200
Qy 1401 GlyGlnIleCySLeuTyPheValThyValProIySaPtyValaVlaPheGlyGlnPhe 1420
Db 4201 GGACAGATCTGCTGTGTAATTTGTACTGTGCCAAAGACTATGTGTGTGGCCAGTTC 4260
Qy 1421 AlaPhePheHieThraIaLeuAnaPyaValaGluValaHiaaPGLyHiaSerGlnHia 1440
Db 4261 GCCTTCCTTTCACACGGCCCTCAACGACGTGGTGGAGTTCAACGCGCACAGCCAGC 4320
Qy 1441 SerArgLeuLeuSerSerLeuSerGlySerHieThraGlyGlySerLeuProLeuAlaThr 1460
Db 4321 TCGGGGCTCTCAAGCTCCCTCTCGGGCTCCATACAGAGAAATCACTGCTTGGCCACC 4380
Qy 1461 SeranGlnValaLeuIleLysPheSerAlaLysGlyLeuAlaProAlaArgGlyPheHia 1480
Db 4381 TCCATCAAGTTCTCATTAAGTTACGGCCAAAGGCTCGCACAGCCAGAGGCTTCAC 4440
Qy 1481 PheValTyArgAlaValaProArgThSerAlaThraGlnCySerSerValaProGluPro 1500
Db 4441 TTTGTCTACCAAGCGGTTCTCTGAAACGCGCACGACGTGAGTGTGTGTGCGGAAACC 4500
Qy 1501 ArgTyGlyLysArgLeuGlySerAspPheSerValaGlyAlaIleValaArgPheGlnCys 1520
Db 4501 CGCTATGGCAAGAGCTGGGAGTGAATTCGTGGGGGCGATGTCGCTTGCATGAGC 4560
Qy 1521 AasnSerGlyTyAlaLeuGlnGlySerProGluIleGluCySLeuProValaProGlyAla 1540
Db 4561 AACTCGGCTATGCCCTGCAGGGGCGCCAGAGATCGAGTGCCTCCGTGCTGGGGCC 4620
Qy 1541 LeuAlaGlnTPaaanaValaSerAlaProThrCySValaValaProCySGLyValaLeuThr 1560
Db 4621 TTGGCCCAATGAAATGTCTCAGCGCCACGTGTGTGGTGGGAGGCAACCTTCACA 4680
Qy 1561 GluArgArgGlyThraIleLeuSerProGlyPheProGluProIyThraaenSerLeuasn 1580
Db 4681 GAGGCGAGGGGACCATCTGCTCCCTGGCTTCCAGAGCGTACCTTAACAGCCTCAAC 4740
Qy 1581 CySValaTPLySileValaValaProGluGlyAlaGlyIleGlnIleGlnValaValaSerPhe 1600
Db 4741 TGTGTGTGGAAGATGTGTGTCGCCGAAGCCCTGGCATCCAGATCCAAAGTTGTCAATT 4800
Qy 1601 ValThraGlnIaasnTPaaanSerLeuGluValaPheaspGlyAlaasnThraValaThr 1620
Db 4801 GTGACAGAGCAAGACTGGGACTCGCTGGAAGTATTGATGTGTGACGATTAACACTGTAAAC 4860
Qy 1621 MetLeuGlySerPheSerGlyThraThraValaProAlaLeuLeuasnSerThraasnGln 1640
Db 4861 ATGCTGGGAGATTCTCAGAAACAACCTGCTGCTCTTGAACAAGACCTCAACACAG 4920
Qy 1641 LeuTyThraLeuHiaPheTySerAspIleSerValaSerAlaIleGlyPheHiaLeuGly 1660
Db 4921 CTCAACCTTCATTTCTACTCAGATATCAGCGTATCTGACGCTTCCACTTGGAGTAC 4980
Qy 1661 LysThraValaGlyLeuSerSerCySProGluProAlaValaProSeranGlyValaLysThr 1680
Db 4981 AAAACGGTGGGCTGAGCAGATTGTCCGGAACCTGCTGTGCCAGTAACGGGGTGAAGACT 5040
Qy 1681 GlyGlyLysArgTyThraLeuValaAsnAspValaValaSerPheGlnCySGLyProGlyTyAlaLeu 1700
Db 5041 GGCGAGCCCTACTGTGAATGATGTGTGTCTTTCAGTGTGAGCCGGGATATGCCCCCTC 5100
Qy 1701 GlnGlyHiaValaHiaIleSerCySweSerProGlyThraValaArgArgTPaaanTyProPro 1720
Db 5101 CAGGGCCAGCCCAATCTCTCGATGCCGGAACAGTGGCGGCGATGAACTACCTCTCT 5160
Qy 1721 ProLeuCySValaIleGlnCySGLyGlyThraValaGluIleuMetGluGlyValaIleLeuSer 1740
Db 5161 CCACCTGTATATGACAGTGTGGGGAACAGTGAAGAGATGAGGGGCTGTATCTGAC 5220
Qy 1741 ProGlyPheProGlyAasnTyProSerAsnMetAspCySserTPLySileAlaLeuPro 1760
Db 5221 CCGGCTTCCAGGCAACTACCCAGTAACAGTACGCTCTCTGAAAAATAGCATGCTGCC 5280
Qy 1761 ValaGlyPheGlyAlaHiaIleGlnPheLeuasnPheSerThraGluProaenHiaSAPTy 1780
Db 5281 GTGGGCTTGGAGCTCAATCCAGTTCCTGAACCTTCCACCGAGCCCAACAGACTAC 5340
Qy 1781 IleGluIleArgAnaGlyProTyGluThraSerArgMetMetGlyArgPheSerGlySer 1800
Db 5341 ATGAATAATCCGGAATGGCCCTTAAGAACAGCGCGATGATGGAAATTCAGTGAAGC 5400
Qy 1801 GluLeuProSerSerLeuLeuSerThraSerHieGluThraThraValaTyPheHiaSerAsp 1820
Db 5401 GAGCTTCCAAAGCTCTCTCTCTCCAGCTCCACAGAACCAACCGTGTATTTCCACAGCAGC 5460
Qy 1821 HisSerGlnaanaArgProGlyPheLysLeuGluTyGlnAlaTyGluLeuGlnCys 1840
Db 5461 CACTCCAGATCGGCGAGATTCAGCTGGAATATCAGCGCTATGAATTCGAAGATGC 5520
Qy 1841 ProaapProGluProPheAlaIleasnGlyIleValaArgGlyValaGlyTyThraaenValaGlyGln 1860
Db 5521 CCAGACCCAGAGCCCTTTCGCAATGGATGTGGAGGAGACTGGCTTCACAGTGGACAA 5580
Qy 1861 SerValaThraPheGluCySLeuProGlyTyGlnLeuThraGlyHiaSerProValaLeuThraCys 1880
Db 5581 TCACTGACTTTCAGATGCTCTCCGGGATATCAATGACTGACCACCTGTCTCACTGT 5640
Qy 1881 GlnHieGlyThraaanaArgaenTPaaanPheProLeuProIyCySGLyValaProCySGLy 1900
Db 5641 CAACATGCAACCAACCGAACTGGGACCAACCCCTGCCAAGTGTGAAGTCCCTTGGGC 5700
Qy 1901 GlyAsnIleThraSerSeranGlyThraValaTySerProGlyPheProSerProTySer 1920
Db 5701 GGGAAACATCACTTCTTCACAGGACGTGTATCTCCCGGGGTCCCTACCCCTACTCC 5760
Qy 1921 SerSerGlnaapCySValaTPLeuIleThraValaProIleGlyHiaGlyValaArgLeuAsn 1940
Db 5761 AGCTCCAGAGACTGTGCTGGCTGATCAACCGTGCCTATGGCCATGGCGCTCCGCTTAAC 5820
Qy 1941 LeuSerLeuLeuGlnThraGluProSerGlyAspPheIleThraIleTPaaapGlyProGln 1960
Db 5821 CTGAGCTGTGACAGACAGAGCCCTGTGAGATTTCATCAACATCTGGATGGGCCACAG 5880
Qy 1961 GlnThraIaProArgLeuGlyValaPheThraArgSerMetAlaLysLysThraValaGlnSer 1980
Db 5881 CAACACACACACAGCGCTCGGCTCTTCACCCGGAGCTGGCCAAAGAAACAGTGCAGACT 5940
Qy 1981 SerSeranGlnValaLeuLeuLysPheHiaArgAspAlaIlaThraGlyGlyIlePheAla 2000
Db 5941 TCATCCAACAGATCTGCTCAAGTTCCACCGGTATCAAGCCACAGCAGAGGGAGATCTTGCC 6000
Qy 2001 IleAlaPheSerAlaTyProLeuThraLysCySProProProThraIleLeuProaenAla 2020
Db 6001 ATAGCTTCTTCGGTTATTCATCTCAACCAATGCCCTCTCCACCAATCTCCCAAGGCC 6060
Qy 2021 GluValaValaThraGluasnGluGluPheAsnIleGlyAspIleValaArgTyThraGlyLeu 2040
Db 6061 GAAGTCTGTCAGAGATGAAGAAATTCAAATATGTGACATCGTACCTCAAGATGCTC 6120
Qy 2041 ProGlyPheThraLeuValaGlyasnGluIleLeuThraCySlysLeuGlyThraTyLeuGln 2060
Db 6121 CTGGCTTTACCTTATGGGGAATGAATTCATCTGCAAACTTGGAACTTACCTGACG 6180
Qy 2061 PheGluGlyProProProIleCySGLyValaHiaCySProThraasnGluLeuThraAsp 2080
Db 6181 TTGGAAGACCAACCCGATATGTGAAGTCACTGTCCAAATATGAGCTTCTGACAGAC 6240
```

QY 2081 SerThrGlyValIleuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCys 2100  
DB 6241 TCACAGGGGTATCTCTGAGCCAGAGCTACCTGGAGAGCTATCCCACTTCCAGACCTGC 6300  
QY 2101 SerTyrPheValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeu 2120  
DB 6301 TCTTGCTGTGGAGAGTGGAGCCGACTATACATCTCCCTCAGAGTGGATCTTCTTC 6360  
QY 2121 SerGluPheGlnTyrPheGlnPheGlnIlePheAspGlyProSerGlyGlnSerProLeu 2140  
DB 6361 ACCGAGAGCAATATAGATGAGTTTGAATTTTGAATGCTCATAGAGACAGAGCTCTG 6420  
QY 2141 LeuValAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerSerAsnSer 2160  
DB 6421 CTGAAGCCCTCAGTGGGAAATTACTCACTCCCTGATTTGACACAGCTCAAGCACTCT 6480  
QY 2161 ValTyrLeuAlaGlyTyrPheSerSerPheAspHisAlaTyrAsnArgGlyPheValIleArgTyr 2180  
DB 6481 GTGTACCTGCGTGTGCTCATCTGATCATCCCTACCAATCGAAGGGCTTCAAGATCGCTAT 6540  
QY 2181 SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr 2200  
DB 6541 TCAGGCCCTTACTGACAGCTTCCAGGAGCTCAGCTCAGTGGCTTCATTCAGGCCAGACC 6600  
QY 2201 SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly 2220  
DB 6601 AGCACCCAGCCGGGGGCTCCATCCATCTTGGCTGCAACGCGGCTACCCCTGTGGGA 6660  
QY 2221 HisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuTyrPheSerGlyAlaIle 2240  
DB 6661 CACAGCATGGCCATCTGTACCCGGGACCCCGAGGGCTTACCACTGTGGAGCGAAGCATC 6720  
QY 2241 ProLeuGlyGlnAlaLeuSerCysGlyLeuProGlyAlaProIleAsnGlyMetValPhe 2260  
DB 6721 CCTCTGTGTAAGCTCTTCTTCTGTGGCTTCTGTAGGCCCCCAAGATGGATGGTGT 6780  
QY 2261 GlyGlySerGlyTyrThrValGlyTyrLysAlaValTyrSerCysSerGlyGlyTyrHisLeu 2280  
DB 6781 GGCAGAGAGTACACAGTGGGAGAACCAAGCGGTGTACAGTGTGAGTGAAGGCTTCCACTC 6840  
QY 2281 GlnAlaGlyAlaGluAlaThrAlaGluCysLeuAspPheThrGlyLeuTyrPheAsnAlaGln 2300  
DB 6841 CAGGACAGGCTGTAGGCGACTGTAGAGATGTCTGTAGACACAGGCTTATAGAGCAACCCCAAT 6900  
QY 2301 ValProProGlnCysValProValThrCysProAspValSerSerIleSerValGluHis 2320  
DB 6901 GTCCCAACACAGTGTGTCTGTGTGCTGTGTGATGTGTGATGTGTGATGTGTGATGTGTG 6960  
QY 2321 GlyArgThrArgLeuIlePheGlyLysThrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIle 2340  
DB 6961 GGCAGATGGAGGCTTATCTTGTAGACACAGTATCAGTTCAGGCGCCAGCTGATGTCTCATC 7020  
QY 2341 CysAspProGlyTyrTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLys 2360  
DB 7021 TGTGACCTGTGCTACTATCTGCTGCAAGAGGTATCTGCTGTGTAGGCGCAATGGCAAA 7080  
QY 2361 TyrSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeuProIle 2380  
DB 7081 TGGAGCTCGGGGAGCTTACGCGCCACCTGCGCAATCATCTCTGTGTGAGAGCTCCGANT 7140  
QY 2381 ProProAsnGlyHisArgIleGlyThrLeuSerValTyrGlyAlaThrAlaIlePheSer 2400  
DB 7141 CCCCCCAATGGCCAGCATCGGAGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7200  
QY 2401 CysAsnSerGlyTyrThrLeuValGlySerArgValArgGluCysMetAlaAsnGlyLeu 2420  
DB 7201 TGGCAATTCGGATACACACATGGGGCTCCAGGGTGTGCTGTGTGTGTGTGTGTGTGTGT 7260  
QY 2421 TyrSerGlySerGlyValArgCysLeuAlaGlyHisCysGlyTyrProGluProIleVal 2440  
DB 7261 TGGAGTGTCTGTGAAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7320  
QY 2441 AsnGlyHisIleAsnGlyGluAsnTyrSerTyrArgGlySerValValTyrGlnCysAsn 2460  
DB 7321 AACGACACATCAATGGGAGAACTACACTACCCGGGAGGTGTGTGTGTGTGTGTGTGTGT 7380  
QY 2461 AlaGlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisIleTyrPhe 2480  
DB 7381 GTTGGCTTCCGCTGTATGT 7440  
QY 2481 GlyLysThrProPheCysValProIleThrCysGlyHisPheProGlyAsnProValAsnGly 2500  
DB 7441 GGCAGAACCCCTTCTGT 7500  
QY 2501 LeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValLysPheValCysAsnProGly 2520  
DB 7501 CTCACTCAGGTTACCAAGTTTAACTTCAACGATGTGTGTGTGTGTGTGTGTGTGTGTGT 7560  
QY 2521 TyrMetAlaGluGlyAlaAlaIleArgSerGlnCysLeuAlaSerGlyGlyTyrPheSerPhe 2540  
DB 7561 TATATGCTGAGAGGGGCTGTAGTCCCAATGCTGTGGCCAGCGGCAATGTAGTACATG 7620  
QY 2541 LeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGluAsnSerValArg 2560  
DB 7621 CTGCCCACTGAGATCATCACTGTACATCTGTGACACCAAGAAATATGTGTGTGT 7680  
QY 2561 GlnValHisAlaSerGlyProHisArgPheSerPheGlyTyrThrValSerTyrArgCys 2580  
DB 7681 CAGGTCCAGCGCAGCGGCCCGCAGAGTTTCAAGTTTCCAGACACTGTGTCTTACCGGTGC 7740  
QY 2581 AsnHisGlyPheTyrLeuLeuGlyTyrProValLeuSerCysGlnGlyAspGlyThrThr 2600  
DB 7741 AACCAAGGCTTACTCTCTGTGGACCCCACTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 7800  
QY 2601 AspArgProArgProGlnCysLeuLeuValSerCysGlyHisPheProGlySerProHis 2620  
DB 7801 GACCGTCCCGCCCAAGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7860  
QY 2621 SerGlnMetSerGlyAspSerTyrThrValGlyAlaValAlaGlyTyrSerCysIleGly 2640  
DB 7861 TCCAGATGTGTGAGACAGTATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7920  
QY 2641 LysArgThrLeuValGlyAsnSerThrArgMetCysGlyLeuAspGlyHisIleTyrThrGly 2660  
DB 7921 AAGCTACTCTGTGTGGGAAACAGCACCGCATGTGTGTGTGTGTGTGTGTGTGTGTGT 7980  
QY 2661 SerLeuProHisCysSerGlyThrSerValGlyValCysGlyAspProGlyTyrProAla 2680  
DB 7981 TCCCTCTCACTGTCTGAGAACCAAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8040  
QY 2681 HisGlyIleArgLeuGlyAspSerPheAspProGlyTyrThrValMetArgPheSerCysGly 2700  
DB 8041 CATGGCATCGCTTGTGGGGAGACGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8100  
QY 2701 AlaGlyHisValLeuArgGlySerSerGluArgThrCysGlnAlaAsnGlySerTyrPhe 2720  
DB 8101 GCTGGCAAGCTGTCCGGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8160  
QY 2721 GlySerGlnProGluCysGlyValIleSerCysGlyAsnProGlyTyrProSerAsnAla 2740  
DB 8161 GGTCTGAGGCTGTAGT 8200  
QY 2741 ArgValAlaPheSerAspGlyLeuValPheSerSerSerIleValTyrGluCysArgGly 2760  
DB 8221 CAGATGTGTGTCAAGTATGT 8280  
QY 2761 GlyTyrTyrAlaThrGlyLeuLeuSerArgHisCysSerValAsnGlyTyrThrTyrGly 2780  
DB 8281 GGAATCTACGCCACAGGCTGT 8340  
QY 2781 SerAspProGluCysLeuValIleAsnCysGlyAspProGlyTyrLeuProIleAsnGlyLeu 2800  
DB 8341 AGTGAACCTGTAGT 8400  
QY 2801 ArgLeuGlyAsnAspPheArgTyrAsnLysThrValThrTyrGlnCysValProGlyTyr 2820

Db 8401 CGCGTGGGCAATGACTTCAGGTAACAACAACTGACATATCACTGTCCTCGGCTAT 8460  
Qy 2821 MetMetGluSerHisArgValSerValLeuSerCyThrLysAspArgThrTTPaangly 2840  
Db 8461 ATGATGGAGTACATAGATGATCTGTCTGAGCTGACCAAGAGACCGGACATGAAATGGA 8520  
Qy 2841 ThrLysProValCysLysAlaLeuMetCysLysProProPoleuLeuProAsnGlyLys 2860  
Db 8521 ACCAAGCCCGCTCTGCAAGCTCTCATGTGCAAGCCACTCCGCTCATCCCAATGGAGAG 8560  
Qy 2861 ValValGlySerAspPheMetTTPGlySerSerValThrTyrAlaCysLeuGluGlyTyr 2880  
Db 8581 GTGGTGGGGTCTGACTTATGAGGGGCTCAAGTGTGACTTATGCTGCTGGAGGGGATAC 8640  
Qy 2881 GluLeuSerLeuProAlaValPheThrCysGluGlyAsnGlySerTTPThGlyGluLeu 2900  
Db 8641 CAGCTCTCCCTGCGCGCGGTGTTCACCTGTAGAGGAAATGGGTCTGACCGGAGAGCTG 8700  
Qy 2901 ProGlnCysPheProValPheCysGlyAspProGlyValProSerArgGlyArgArgGlu 2920  
Db 8701 CCTCAGTGTTCCTGCTGTCTGCGGGATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8760  
Qy 2921 AspArgGlyPheSerTyrArgSerSerValSerPheSerCysHisProProLeuValLeu 2940  
Db 8761 GACCGAGGCTCTCTCTACAGGTCACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 8820  
Qy 2941 ValGlySerProAlaArgArgPheCysGlnSerAspGlyThrTTPSerGlyThGlnProSer 2960  
Db 8821 GTGGGCTCTCCACGACGAGTTTGGCCAGTGGACATGGAGATGAGTGGACCCAGCCAGC 8880  
Qy 2961 CysIleAspProThrLeuThrThrCysAlaAspProGlyValProGlnPheGlyTlGln 2980  
Db 8881 TGCAATAGTCCGACCTTACCACTGTCGAGACCTGTGTCGACACATTTGGATACAG 8940  
Qy 2981 AsnAsnSerGlnGlyTyrGlnValGlySerThrValLeuPheArgCysGlnLysGlyTyr 3000  
Db 8941 AACATTCTCAGGGGTACCAAGTGGAGACAGTCCCTCCGTTGCAAAAAGGCTAC 9000  
Qy 3001 LeuLeuGlnGlySerThrThrArgThrCysLeuProAsnLeuThrTTPSerGlyThrPro 3020  
Db 9001 CTGCTTCCAGGGCTCCACCAAGGACCTGCTCCCAACCTGACCTGAGTGGAAACCCCA 9060  
Qy 3021 ProAspCysValProHisHisCysArgGlnProGluThrProThrHisAlaAsnValGly 3040  
Db 9061 CTTAGCTGTGTCCTCCACACTGACGAGGAGGACGACCAAGCATGCCAAGCTGGAG 9120  
Qy 3041 AlaLeuAspLeuProSerMetGlyTyrThrLeuIleThrProAlaArgArgAlaSerPro 3060  
Db 9121 GCCCTGATTTGGCTTCATGGGCTACACGCTCATTACTCTGACAGAGGGCTTCTCC 9180  
Qy 3061 SerArgValAlaProSerThrAlaProAlaArgMetAlaAlaGlyAlaSerArg 3080  
Db 9181 TCAAGGGGTGCTCCGACGACCGCACTGTCAAGGCGAGGACGCTGGACAGCAAGCGCG 9240  
Qy 3081 ProSerIleATPArgSerGlyProValGlyAspProSerThrLeuProGlySerHisArg 3100  
Db 9241 CCACTGCTGAGAGTCCGCGCCAGTGGAGAACCCATCAACATGCGCGGAGCCACCGC 9300  
Qy 3101 SerProLysPro 3104  
Db 9301 TCACCCAAAGCT 9312

RESULT 4

ADH71167  
ID ADH71167 standard; DNA; 10136 BP.

XX AC ADH71167;

XX DT 25-MAR-2004 (first entry)

XX DE Human gene of the invention NOV4q SEQ ID NO:63.

Kw ds; gene; human; cytosolic; immunomodulator; neuroprotective; nootropic;  
Kw anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;  
Kw vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;  
Kw obesity; diabetes; infectious disease; metabolic syndrome X;  
Kw dyslipidaemia.  
OS Homo sapiens.  
XX PN WO2003102155-A2.  
XX PD 11-DEC-2003.  
XX PF 03-JUN-2003; 2003WO-US017430.  
XX PR 03-JUN-2002; 2002US-0385120P.  
PR 04-JUN-2002; 2002US-0385784P.  
PR 05-JUN-2002; 2002US-0386041P.  
PR 05-JUN-2002; 2002US-0386047P.  
PR 06-JUN-2002; 2002US-0386376P.  
PR 06-JUN-2002; 2002US-0386453P.  
PR 06-JUN-2002; 2002US-0386854P.  
PR 06-JUN-2002; 2002US-0387016P.  
PR 07-JUN-2002; 2002US-0387766P.  
PR 07-JUN-2002; 2002US-0386816P.  
PR 07-JUN-2002; 2002US-0386931P.  
PR 07-JUN-2002; 2002US-0386942P.  
PR 07-JUN-2002; 2002US-0386971P.  
PR 07-JUN-2002; 2002US-0387262P.  
PR 08-JUN-2002; 2002US-0296960P.  
PR 10-JUN-2002; 2002US-0387400P.  
PR 10-JUN-2002; 2002US-0387535P.  
PR 11-JUN-2002; 2002US-0387610P.  
PR 11-JUN-2002; 2002US-0387625P.  
PR 11-JUN-2002; 2002US-0387634P.  
PR 11-JUN-2002; 2002US-0387668P.  
PR 11-JUN-2002; 2002US-0387696P.  
PR 11-JUN-2002; 2002US-0387702P.  
PR 11-JUN-2002; 2002US-0387836P.  
PR 11-JUN-2002; 2002US-0387859P.  
PR 12-JUN-2002; 2002US-0387933P.  
PR 12-JUN-2002; 2002US-0387934P.  
PR 12-JUN-2002; 2002US-0387960P.  
PR 12-JUN-2002; 2002US-0388022P.  
PR 12-JUN-2002; 2002US-0388066P.  
PR 13-JUN-2002; 2002US-0389123P.  
PR 14-JUN-2002; 2002US-0389118P.  
PR 14-JUN-2002; 2002US-0389120P.  
PR 14-JUN-2002; 2002US-0389144P.  
PR 14-JUN-2002; 2002US-0389146P.  
PR 17-JUN-2002; 2002US-0389729P.  
PR 17-JUN-2002; 2002US-0389742P.  
PR 18-JUN-2002; 2002US-0389884P.  
PR 19-JUN-2002; 2002US-0390006P.  
PR 19-JUN-2002; 2002US-0390209P.  
PR 21-JUN-2002; 2002US-0390763P.  
PR 21-JUN-2002; 2002US-0390766P.  
PR 06-AUG-2002; 2002US-0401628P.  
PR 09-AUG-2002; 2002US-0402156P.  
PR 09-AUG-2002; 2002US-0402256P.  
PR 09-AUG-2002; 2002US-0402389P.  
PR 12-AUG-2002; 2002US-0402786P.  
PR 12-AUG-2002; 2002US-0402816P.  
PR 12-AUG-2002; 2002US-0402819P.  
PR 12-AUG-2002; 2002US-0402832P.  
PR 13-AUG-2002; 2002US-0403448P.  
PR 13-AUG-2002; 2002US-0403459P.  
PR 13-AUG-2002; 2002US-0403531P.  
PR 13-AUG-2002; 2002US-0403532P.  
PR 13-AUG-2002; 2002US-0403563P.  
PR 13-AUG-2002; 2002US-0403617P.  
PR 15-AUG-2002; 2002US-0403617P.  
PR 26-AUG-2002; 2002US-0406182P.  
PR 26-AUG-2002; 2002US-0406355P.



QY 401 AAlaProValIleuGIYThrPheSerGIYAsnGluProSerSerIleThrSerSerGIY 420  
Db 1201 GGGCCGCTCTGGGCACTTCTTCAGAAACAGCTTCTCTTCATACAAAGAGTGC 1260  
QY 421 HlValAlaArgLeuGIYpHeGlnThrAspHisSerThrGIYLyAsrGIYpHeAsnIle 440  
Db 1261 CACGAGGCCCGCTCGAGTTCAGACTGACCACTCCACAGGGAAGAGGGCTTCAACATC 1320  
QY 441 ThrPheThrThrPheArgHisAsnGluCysProAspProGIYValProValAsnGIYLyS 460  
Db 1321 ACTTTTACCACCTTCCGACACAAACAGTGCCTGGGCTTCCAGTAAATGGCAAA 1380  
QY 461 ArgPheGIYAspSerLeuGluLeuGIYSerSerIleSerPheLeuCysAspGluGIYpHe 480  
Db 1361 CGGTTGGGGAGACGCTCCAGCTGGGAGCTTCATCTCTCTCTGTGATGAAGGCTTC 1440  
QY 481 LeuGIYThrGIYArgIYSerGIYThrIleThrCysValIleuLySGluGIYSerValIleTrp 500  
Db 1441 CTGGGACTCAGGGCTCAGAGACCATCACTGCGTCCGGAAGAGGAGCAGCTGCTG 1500  
QY 501 AsnSerAlaValIleuArgCysGluAlaProCysGIYLyHisLeuThrSerProSerGIY 520  
Db 1501 AACAGCGCTGTGCTGGGTGTGAAGCTCCGTGGTGTCACTGACTTCCGCCAGCGGC 1560  
QY 521 ThrIleLeuSerProGIYTrpProGIYpHeTrpLyAspAlaLeuSerCysAlaIleTrpVal 540  
Db 1561 ACCATCTCTCTCCGGGCTGGCTGTGCTTTCACAAAGATGCTTGAGTGTGGCTGGGTG 1620  
QY 541 IleGluAlaGlnProGIYTrpProIleLyHisIleThrPheAspArgPheLeuThrGIYVal 560  
Db 1621 ATTGAGGCCCAAGCCAGCAGGCTACCCCATCAAAATCACCTTCGACAGATTCAAAACGAGGT 1680  
QY 561 AsnTrpAspThrLeuGluValArgAspGIYArgThrTrpSerAlaProLeuIleGIYVal 580  
Db 1681 AACATAGACCCCTGGAGAGTACGCGATGGGCGGACTTACTCAGCGCCCTTATCGGGGT 1740  
QY 581 TyrHisGIYThrGlnValProGlnPheLeuIleSerThrSerAsnTrpLeuTrpLeuLeu 600  
Db 1741 TACACGGGAGCCCAAGATCCCAAGTCTCTCATCAGACACCAACATCTTACTCTCTC 1800  
QY 601 PheSerThrAspLySerSerHisSerAspIleGIYpHeGlnLeuArgTrpGIYThrIleThr 620  
Db 1801 TTCTCTACCAAGACAGAGTCACTCGGACATCGGCTTCCAGCTCCGTAAGACATTAACA 1860  
QY 621 LeuGlnSerAspHisCysLeuAspProGIYIleProValAsnGIYGlnArgHisGIYAsn 640  
Db 1861 CTGCAGTCAAGACCACTGTCTGGATCCAGAAATCCAGTAAATGACAGCGTCAATGGGAT 1920  
QY 641 AspPheTrpValGIYAlaLeuValThrPheSerCysAspSerGIYTrpThrLeuSerAsp 660  
Db 1921 GACTTCTACGTTGGGGCGCTGATGACCTTCAGCTGTGACTGGGCTTACACATTAAGTAC 1980  
QY 661 GIYGIYProLeuGluCysGIYProAsnPheGlnTrpSerArgAlaLeuProSerCysGIY 680  
Db 1981 GGGAGAGCTCTGAGAGTGTGAGCCCACTTCAAGTGAAGCCGGGCTTCCAGTTGTGAA 2040  
QY 681 AlaLeuCysGIYGIYpHeIleGlnGIYSerSerGIYThrIleLeuSerProGIYpHePro 700  
Db 2041 GCTCTGTGTGGCTTCATTCAAGGCTCCAGTGGAGACATCTTGTGCCAGGGTTCCT 2100  
QY 701 AspPheTrpProAsnAsnLeuAsnCysThrTrpIleIleGIYThrSerHisGIYLyGIY 720  
Db 2101 GACTTCTACCCCAACACTTGAACCTGACCTGATATACGAAACATCTCAAGGAGAGGT 2160  
QY 721 ValIlePheThrThrPheHisThrPheHisLeuGIYSerGIYHisAspTrpLeuLeuIleThr 740  
Db 2161 GTGTCTTCACTTCCACACCTTCCAGCTGAAAAGTGGCCATGACTACTCTTACTACT 2220  
QY 741 GluAsnGIYSerPheThrGlnProLeuArgGlnLeuThrGIYSerArgLeuProAlaPro 760  
Db 2221 GAGAAAGGAGCACTTCAACCAAGCCCTGAGGAGCTAACTGATCTCGGCTCCAGCTCCC 2280

QY 761 IleSerAlaGluLeuTrpGIYAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780  
Db 2281 ATCAGCGCTGGGCTCTTATGGCACTTCACTCCCAAGGTCGCTTCAATCTGATTTCTCC 2340  
QY 781 MetSerTrpGIYGIYpHeAsnIleThrPheSerGIYTrpAspLeuGluProCysGIYGIY 800  
Db 2341 ATGTCAATGAAGATTCACATCACTTTCAGAGTACGACTTGGAGCCCTGTGAGAGAG 2400  
QY 801 ProGluValProAlaTrpSerIleArgLySGluLeuGlnPheGIYValGIYAspThrLeu 820  
Db 2401 CCCAGAGTCCAGGCTTACAGCATCCGGAAGGGCTTGGAGTTGGCTGGGCGCAACCTTG 2460  
QY 821 ThrPheSerCysPheProGIYTrpArgLeuGluGIYThrAlaArgIleThrCysLeuGIY 840  
Db 2461 ACCTTCTCTCTCTCCCGGCTACCGTGTGAGGAGCACGCCCATCACGCTGCGGG 2520  
QY 841 GIYArgArgArgLeuTrpSerSerProLeuProArgCysValAlaGIYCysGIYAsnSer 860  
Db 2521 GGCAAGCGGGCTGTGGAGCTCGCTCTGCAAGTGTGTGTGAGTGGGAAATTCA 2580  
QY 861 ValIleThrGIYThrGIYGIYThrLeuLeuSerProAsnPheProValAsnTrpAsnAsn 880  
Db 2581 GTCAAGGCACTCAGGGTACTTGTCTGCCCCCACTTCTCTGGAACATAAACAAT 2640  
QY 881 HisGluCysIleTrpSerIleGlnThrGlnProGIYLySGlyIleGlnLeuLyAlaArg 900  
Db 2641 CATGAATGATCTATCTCCATCCAGACCCAGCAGGGAAGGGAATTCAGCTGAAGCCAGG 2700  
QY 901 AlaPheGluLeuSerGIYGIYAspValIleuLyValTrpAspGIYAsnAsnAsnSerAla 920  
Db 2701 GCATTTCGAACCTTCGAAGAGAGATGCTCTCAAGTTTATGATGGCAACACAACTCCGCC 2760  
QY 921 ArgLeuLeuGIYValPheSerHisSerGIYMetMetGIYValIleLeuAsnSerThrSer 940  
Db 2761 CGTTGCTGGAGTGTTTAGGCATTTGAGATGTAGGGGGTGACTTTGAACAGACATCC 2820  
QY 941 SerSerLeuTrpLeuAspPheIleThrAspAlaGluAsnThrSerLySGlyPheGluLeu 960  
Db 2821 AGCAGCTGTGGCTGATTTCAATCATCACTGATGCTGAACACCAAGAGGCTTTGAACGTG 2880  
QY 961 HisPheSerSerPheGluLeuIleLyCysGIYAspProGIYTrpProLyPheGIYTrp 980  
Db 2881 CACTTTCGAAGCTTGAACATCATCAATGTGAGAGACCAAGAAACCCCAAGTTTGGCTAC 2940  
QY 981 LyValHisAspGIYGIYHisPheAlaGIYSerSerValSerPheSerCysAspProGIY 1000  
Db 2941 AAGGTTCAATGATGAAGGTCAATTTTGCAGGAGGCTCCGTGCCCTTCACTGACCTGGA 3000  
QY 1001 TyrSerLeuArgGIYSerGIYGIYLeuLeuCysIleLeuSerGIYGluArgArgThrTrpAsp 1020  
Db 3001 TACAGCTGTGGGTTAGTGAAGAGCTGTGTGTGAGTGAAGAGCCCGGACCTTGGGAC 3060  
QY 1021 ArgProLeuProThrCysValAlaGIYCysGIYGIYThrValArgGIYGIYValIleSerGIY 1040  
Db 3061 CGGCTCTGCCCACTGTGTGCCGAGTGTGAGAGGACATGAGAGAGAGAGTGTCCGGG 3120  
QY 1041 GlnValLeuSerProGIYTrpProAlaProTrpGIYHisAsnLeuAsnCysIleTrpThr 1060  
Db 3121 CAGGTGTGTACACCGGGGTATCCAGCTCCCATATGAACAAATCTCACTCATCTGAGAC 3180  
QY 1061 IleGluAlaGluAlaGIYCysThrIleGIYLeuHisPheIleValIleAspThrGluGlu 1080  
Db 3181 ATCGAAGCAGAGCGGCTCACCACATTTGGGCTACATTCCTGGTGTGACACAGAGAGAG 3240  
QY 1081 ValHisAspValIleuArgIleTrpAspGIYProValGluSerGIYValIleLeuLyGIY 1100  
Db 3241 GTTCACAGCACTGTGCGCATCTGGAGTGGGCTGTGAGAGCGGGGTTCTGTGAGAGAG 3300  
QY 1101 LeuSerGIYProAlaLeuProLyAspLeuHisSerThrPheAsnSerValIleLeuGln 1120  
Db 3301 CTGAGTGGCCCGGCTCGCCCAAGAGACCTGCATAGACCTTCAACTGCGTGTGCTGAG 3360  
QY 1121 PheSerThrAspPheThrSerLySGlnGIYPheAlaIleGlnPheSerValSerThr 1140

```
Db 3361 TTGAGCAGTCACTTCTTCAACAGCAGGCGCTTGGCATTCAATTTTCAGTGTCCACA 3420
Qy 1141 AATThSerCySaAsnAspProGly11eProGlnAanglySerArgSerGlyAAspSerTrp 1160
Db 3421 GCACATCTCTGCAATACCTCGAGATCCCGCAGAAATGGGAGTCGAGTGTGACAGTTGG 3480
Qy 1161 GUAAGAGlyAAspSerThrValPheGlnCySaAspProGly11eProGlnAanglySerAla 1180
Db 3481 GAAGCGGCGGACTCCACAGTGTTCAGTGTGACCTCGCTACGCGCTGACAGGAGATGCA 3540
Qy 1181 GUA11eSerCySaVal11eGlnAanglyPhePheTrpGlnProSerProProThCys 1200
Db 3541 GAGATCAGCTGTGTGAAAGATCGAGAACAGGTTCTTCTGCGACGCCAGCCCAACATGC 3600
Qy 1201 11eAlaProCyGly11eGlnAanglyPhePheTrpGlnProSerProProThCys 1220
Db 3601 ATCGCTCCCTGCGGGGAGACTGACAGGACCATCGAGTCATCTCTCACAAATTAC 3660
Qy 1221 ProGlnProTrpProProGly11eGlnCySaAspTrpVal11eThVal11eSerProAspTrp 1240
Db 3661 CCAGAACCTTACCCGCGCAGGAGAGTGTGACCTGGAAGTACCGTCTCCACAGACTAC 3720
Qy 1241 Val11eAla11eVal11ePheAsn11ePheAsnLeuGlnProGly11eYAspPheLeu11e 1260
Db 3721 GTCATCGCTCGGATTTTACATCTTTAACCTGAGGCTGAGCTGATCTCTCATATC 3780
Qy 1261 TyrAspGlyArgAspSerLeuSerProLeu11eGlySerPheTrpGlySerGlnLeuPro 1280
Db 3781 TACAGCGGACGGGACTCTCTACGCCCTTCTATGAAAGCTTCTATGAGCTCCAGCTCCA 3840
Qy 1281 G1yArg11eGlnSerSerSerAsnSerLeuPheLeu11ePheArgSerAspAlaSerVal 1300
Db 3841 GGCAGCATTTGAAACAGCAGCAACAGCCTTCTCTCCCTTCCGACGATGATCTGTC 3900
Qy 1301 SerAsn11eGlyPheVal11eAspTrpThGlnAsnProArgGlnSerCySaPheAspPro 1320
Db 3901 AGCATGCTGAGCTTCTGATTTGATGACTATACAGAAAAACCGGAGTCACTTTTGTATCT 3960
Qy 1321 GlySer11eGlySaanglyThrArgVal11eGlySerAspLeu11eGlySerSerVal11e 1340
Db 3961 GGTTCATCAAGAACAGCAGCAACGCGGTGGGTCCGACCTGAACTGGCTCTCCGTCAC 4020
Qy 1341 TyrTrpCySaHsGlyGlyTrpGlnVal11eGlnGlyThSerThrLeuSerCySa11eLeuGly 1360
Db 4021 TACTACTGCCACGCGGGCTACGAAGTTGAGGGCACCTCGACCTGAGCTGCATCTGGGG 4080
Qy 1361 ProAspGlyLySProVal11eTPAsnAspProArgProVal11eCySaThrAlaProCySaGly 1380
Db 4081 CCGATGCGAAGCCCGTGTGAAACATCCCGGACGAGTGCACAGCCCCCTGTGGGGGA 4140
Qy 1381 GlnTrpVal11eGlySerAspGlyVal11eLeuSerProAsnTrpProGlnAanglyThrSer 1400
Db 4141 CAGATGCTGGGTTCGAGCGAGTGTCTGTCCCAACATACCCCAAGCAACATCAACGCT 4200
Qy 1401 G1yGln11eCySaLeuTrpPheVal11eThVal11eProVal11eAspTrpVal11eGlnPhe 1420
Db 4201 GGAACATGCTGCTGATTTTGTACTGTGCCCCAAGAGCTATGTGTGTGGCCAGTTC 4260
Qy 1421 AlaPhePheH1eThrAlaLeuAsnAspVal11eGlnVal11eAspGlyH1eSerGlnH1e 1440
Db 4261 GCCTTCTTTCACAGGCGCTCAACGAGCGTGTGAGTTCACAGCGCCACAGCCAGC 4320
Qy 1441 SerArgLeuLeuSerSerLeuSerGlySerH1eThrGlyGlnSerLeuProLeuAlaThr 1460
Db 4321 TCGCGGCTCTCAGCTCCCTCTCGGGCTCCATACAGAGAAATCACTGCTTGGCCACC 4380
Qy 1461 SerAsnGlnValLeu11eLySPheSerAla11eGlyLeuAlaProAlaArgGlyPheH1e 1480
Db 4381 TCCAAATCAAGTTCTATTAAAGTTACGCGCAAGGCTGCGCACGCGGAGGCTTCCAC 4440
Qy 1481 PheVal11eGlnAlaVal11eProArgThrSerAlaThrGlnCySaSerSerVal11eProGlnPro 1500
Db 4441 TTGTCTAACCAAGCGGTTCTCGAACCAAGCCAGCAGTGCAGTCTGTGCGGAAACC 4500
Qy 1501 ArgTrpGlyLySaArgLeuGlySerAspPheSerVal11eAla11eArgPheGlnCySa 1520
Db 4501 CGCTATGCGAAGAGCTGTGGAGTGTCTCTGTGGGGGCGCACTGCTCGTTGCAATGC 4560
Qy 1521 AsnSerGly11eVal11eLeuGlnGlySerProGln11eGlnCySaLeuProVal11eProGlyAla 1540
Db 4561 AACTCCGCTATGCTCCCTGACAGGGGTCCGAGAGATGAGTGTCTCTGTGCGGGGCC 4620
Qy 1541 LeuAlaGlnThrAsnVal11eSerAlaProThrCySaVal11eProCyGlyGlyAsnLeuThr 1560
Db 4621 TTGGCCCAATGGAATGTCTCAGCGCCACGTTGTGTGCTGTGTGAGGCACTTACA 4680
Qy 1561 GUAArgArgGly11eThr11eLeuSerProGlyPhePheProGly11eProGlyLeuAsnSerLeuSn 1580
Db 4681 GAGCGGAGGGGACCATCTGTCTCCCTGCTTCCAGAGCCGTACTTCAACAGCTTCAAC 4740
Qy 1581 CySaVal11eTrpVal11eVal11eProGlnGly11eGln11eGlnVal11eSerPhe 1600
Db 4741 TGTGTGTGAAATGATGTGTGTCCCGAAGCGGTGCATCCAGATCAAGTTGTGATTT 4800
Qy 1601 Val11eThGlnGlnAsnTrpAspSerLeuGlnVal11ePheAspGlyAlaAspAsnThrVal11e 1620
Db 4801 GTGACAGACAGAACTGGGACTCGCTGGAAGTATTTGATGTGAGATTAACATGTAACT 4860
Qy 1621 MetLeuGlySerPheSerGly11eThrVal11eProAlaLeuLeuAsnSerThSerAsnGln 1640
Db 4861 ATGCTGGGAGTTTCTCAGGAACAACTGCTGCTGCTTCCAGACAGCACTCCAACTC 4920
Qy 1641 LeuTrpLeuH1eAspTrpSerAsp11eSerAla11eAla11eGlyPheH1eLeuGlnTrp 1660
Db 4921 CTCTACTCTCATTTCTACTCAGATATACGCTATCTCAGCTGCTTCCACTTGGAGTAC 4980
Qy 1661 11eThrVal11eGlyLeuSerSerCySaProGlnProAlaVal11eProSerAsnGlyVal11eTrp 1680
Db 4981 AAAAGGTGGGCTCGAGAGATTGCCGAACCTCTGTGCCCAATACGGGGTAAAGACT 5040
Qy 1681 G1yGlnAanglyTrpLeuVal11eAspVal11eSerPheGlnCySaGlnProGly11eTrpAlaLeu 1700
Db 5041 GCGAGCGCTTCTGTGAAAGATGTGTGTCTTCCAGTGTGAGCGGAGTATGCTTC 5100
Qy 1701 GlnGlyH1eAlaH1e11eSerCySaMetProGly11eThrVal11eArgTrpAsnTrpProPro 1720
Db 5101 CAGGGCAGCGCCACATCTCTGCTGATCCCGGAACAGTGGGAGTAACTACCTCTCT 5160
Qy 1721 ProLeuCySa11eAlaGlnCySaGly11eThrVal11eGlnGlnMetGlnGlyVal11eLeuSer 1740
Db 5161 CCACCTGTGATGCAAGTGTGGGGGAACAGTGGAGAGATGGAGGGCGTGAATCTGTAGC 5220
Qy 1741 ProGlyPheProGlyAsnTrpProSerAsnMetAspCySaSerTrpVal11eAlaLeuPro 1760
Db 5221 CCGGCTTCCAGGCAACTACCCAGTAAACATGAGTGCCTCGGAAAAATGACACTGCCC 5280
Qy 1761 ValGlyPheGlyAlaH1e11eGlnPheLeuAsnPheSerThrGlnProAsnH1eAspTrp 1780
Db 5281 GTGGGCTTGTGAGTGCATCACTGCTGTAATCTTCCACCGACCCCAACACGACTAC 5340
Qy 1781 11eGln11eArgAanglyProGlyTrpThrSerArgMetMetGlyVal11eArgPheSerGlySer 1800
Db 5341 ATGAAATTCGGAGATGGCCCTTATGAGACCGAGCCGATGAGTGGAGATTCACTGAGAAC 5400
Qy 1801 GlnLeuProSerSerLeuLeuSerThrSerH1eGlnThrVal11eTrpPheH1eSerAsp 1820
Db 5401 GAGCTTCCAAAGCTCCCTCTCTCAGCTGCCAGAAACACCGTGTATTTCCACAGCGAC 5460
Qy 1821 H1eSerGlnAsnArgProGlyPheLySaLeuGlnTrpGlnAla11eTrpGlnLeuGlnGlnCySa 1840
Db 5461 CACTCCAGATGCGCAGGATTAAGTGAAGTCAAGGCTTAAGAACTTCAAGAGTGC 5520
Qy 1841 ProAspProGlnProPheAlaAangly11eVal11eArgGlyVal11eGlyTrpAsnVal11eGln 1860
Db 5521 CCAGCCCAAGGCTTGTCCAAATGCAATGTGAGGGGAGCTGAGTCAAACTGGGACAA 5580
```

QY 1861 SerValThrPheGluCysLeuProGlyThrGlnLeuThrGlyHisProValLeuThrCys 1880  
DB TCAGTGCCTTCGAGTGCCTCCGGGATCAATTGACTGGCCACCTGCTCTACGCTG 5640  
QY 1881 GlnHisGlyThrAsnArgAsnTrpAspHisAspLeuProLysCysGluValProCysGly 1900  
DB CAAATGGACCAACCGGAATGGGACCACCCCCTGCCAAGTGTGAAGTCCCTTGAGGC 5700  
QY 1901 GlyAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyrSer 1920  
DB GGGAAACATCACTTCTCCACGGGCACTGTACTCCCGGGGTTCCCTAGCCCTACTCC 5760  
QY 1921 SerSerGlnAspCysValIleTrpLeuIleThrValProIleGlyHisGlyValArgLeuAsn 1940  
DB AGCTCCAGAGCTGTGTGCTGTATATACCGTGCCTATGGCCATGGCGTCCGCTCAAC 5820  
QY 1941 LeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTrpAspGlyProGln 1960  
DB CTCAGCTGCTGCAGACAGAGCCCTCTGGAGATTTCACCATCTGGAGTGGGCCACAG 5880  
QY 1961 GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLysLysThrValGlnSer 1980  
DB CAAACAGACACACGGCTCGGCGCTTCACCCGGAGCATGGCCAGAAACAGTGCAGAGT 5940  
QY 1981 SerSerAsnGlnValIleLeuLeuLysPheHisArgAspAlaAlaThrGlyGlyIlePheAla 2000  
DB TCATCCCAACCAAGGCTGCTCAAGTTCACCGTATCGACGACAGGGGGATCTTGCC 6000  
QY 2001 IleAlaPheSerAlaTyrProLeuThrLysCysAspProProThrIleLeuProAsnAla 2020  
DB ATAGCTTCTCCGCTTATCCATCAACCAAGCCCTCTCCACACATCTCCCAAGCC 6060  
QY 2021 GluValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrArgCysLeu 2040  
DB GAACTCGTCAAGAGATTGAAGATTCAATATAGTGCATCGTACGCTACAGATGCTCC 6120  
QY 2041 ProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuGlyThrTyrLeuGln 2060  
DB CTGGCTTTCCTTAGTGGGAAATGAAATTGACTGCACTGGCAACTTGGAACTTACTCAG 6180  
QY 2061 PheGluGlyProProProIleCysGlyValHisCysProThrAsnGluLeuLeuThrAsp 2080  
DB TTGAAAGGACCAACCCCGATATGTGAGTGCATCTCCACCAATGAGCTTCTACAGAC 6240  
QY 2081 SerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCys 2100  
DB TCACAGGCGGTGATCTGAGCCAGACGATACCTCGMAAGCTATCCCAAGTCCAGACCTGC 6300  
QY 2101 SerTrpLeuValAlaGlyAlaGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeu 2120  
DB TCTTGGCTGTGAGAGTGGAGCCGACATATCACTCCCTCACAGTGGATACTTCTCCTC 6360  
QY 2121 SerGluLysGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeu 2140  
DB AGCAGAAAGCAATATGATGATTTGAATTTTATAGTCCATACGAGACAGATCCTCTG 6420  
QY 2141 LeuLysAlaLeuSerGlyAsnTyrSerAlaProLeuIleValIleThrSerSerAsnSer 2160  
DB CTGAAAGCCCTCAATGGGAATTACTCAGCTCCCTGATTTGCACAGCTCAAGCAACTCT 6480  
QY 2161 ValTyrLeuAspGlyTrpSerSerAspHisAlaTyrAsnArgLysGlyPheLysIleArgTyr 2180  
DB GTTAACTGTGCTTGTGATCTGATCAAGCTTACATCGGAAGGCTTCAATATCCGCTAT 6540  
QY 2181 SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr 2200  
DB TCAAGCCCTTACTGAGCTTCCAGGAGCTCCACTCAAGCTTCAATCTTAGGCAAGAC 6600  
QY 2201 SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly 2220  
DB AGCAGCCAGCCCGGGGCTCATCACTTTGGCTGCAACGCGGCTACCCGCTGTGGGA 6660

QY 2221 HisSerMetAlaIleCysThrArgHisAspProGlnGlyTyrHisLeuTrpSerGluAlaIle 2240  
DB CACAGCATGGCCATCTGTATCCGGCACCCCAAGGCTTACCACTGTGGAGCAAGCATC 6720  
QY 2241 ProLeuCysGlnAlaLeuSerCysGlyLeuProGlnAlaProLysAsnGlyMetValPhe 2260  
DB CTTCTGTCAAGCTTTCTCTGTGGCTTCTGAGGCCCCCAAGATGAAATGATGTTT 6780  
QY 2261 GlyLysGlyThrValGlyThrLysAlaValTyrSerCysSerGlyGlyTyrHisLeu 2280  
DB GGCAGAGAGTACACAGTGGGAACCAAGCCGTGTACAGCTGAGTGAAGCTTCACTC 6840  
QY 2281 GlnAlaGlyValIleGluAlaThrAlaGluCysLeuAspThrGlyLeuTrpSerAsnArgAsn 2300  
DB CAGGCAAGCCCTAGGCACTGCAGAGTGTCTGACACAGGCTTATGAGACCAACCGCAT 6900  
QY 2301 ValProProGlnCysValProValIleThrCysProAspLysSerIleSerValGlnHis 2320  
DB GTCCACACAGTGTGCTGCTGTGACTTGTCTGATGTGATGATGATGATGATGATGAT 6960  
QY 2321 GlyArgTrpArgLeuIlePheGluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIle 2340  
DB GGCAGATGAGGCTTATCTTTGAGACACACTATCAGTTCCAGGCCACGCTGATGCTATC 7020  
QY 2341 CysAspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLys 2360  
DB TGTGACCTTGCTATCTATCTATCTGACCAAGGATCATCCGCTGTCAGGCCAATGGCAA 7080  
QY 2361 TrpSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyLysLeuProIle 2380  
DB TGGAGCTCGGGGCTTACGCCCACTGCGAATATCTCTGTGAGAGCTCCGATT 7140  
QY 2381 ProProAsnGlyHisArgIleGlyThrLeuSerValTyrGlyAlaThrAlaIlePheSer 2400  
DB CCCCCCATGGCCACCGCATCGGAACACTGTCTGTACAGGGGCAACAGCCATCTTCTC 7200  
QY 2401 CysAsnSerGlyTyrThrTrpLeuValGlySerArgValArgGluCysMetAlaAsnGlyLeu 2420  
DB TGCATTCGGATACACACTGTGGCTCCAGGGTGGCTGATGTGATGTGATGTGATGTG 7260  
QY 2421 TrpSerGlySerGluValArgCysLeuAlaGlyHisCysGlyThrProGluProIleVal 2440  
DB TGGAGTGGCTGTGAAGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7320  
QY 2441 AsnGlyHisIleAsnGlyGluAsnTyrSerTyrArgGlySerValValTyrGlnCysAsn 2460  
DB AACGACACATCAATGGGAGAACTACAGCTACCGGGGAGTGTGTGATGCAATGCAAT 7380  
QY 2461 AlaGlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisHisTrpSer 2480  
DB GCTGGCTTCGCTGATCGGATGTCTGTGCGCATCTGCGACAGGATCACTGAGTGC 7440  
QY 2481 GlyLysThrProPheCysValProIleThrCysGlyHisAspProGlyAsnProValAsnGly 2500  
DB GGCAGAGCCCTTCTGTGTGCAATTAACCTGTGAGACACCAAGCAACCTGTCAAGGC 7500  
QY 2501 LeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValIleLysPheValCysAsnProGly 2520  
DB CTCACTCAGGGTAAACAGTTTAACTCAAGATGTGCTCAAGTTTGTTCACACCTCGGG 7560  
QY 2521 TyrMetAlaGluGlyAlaAlaArgSerGlnCysLeuAlaSerGlyGlnTrpSerAspMet 2540  
DB TATATGGCTAAGGGGCTGTAGGTCCCAATGCTGCGCACGGGCAATGAGTCAATG 7620  
QY 2541 LeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGluAsnSerValArg 2560  
DB CTGCCCACTCGAGAAATCACTGTACAGATCTGTGACACCAAGAAATATGAGTTCGT 7680  
QY 2561 GlnValHisAlaSerGlyProHisArgPheSerPheGlyThrThrValSerTyrArgCys 2580  
DB CAGGTCCACCCAGCGGCCGCAAGGTTCAAGTTCGGCACCACTGTGTTCACCGGTGC 7740  
QY 2581 AsnHisGlyPheThrLeuLeuGlyThrProValLeuSerCysGlnGlyAspGlyThrTrp 2600

```
Db 7741 AACCAAGGGCTTACTCTCTGGGACCCAGAGCTTACCTGACAGAGAGATGGCACTAGG 7800
Qy 2601 AAspArgProArgProGlnCysLeuLeuValSerCysGlyValProGlySerProProHis 2620
Db 7801 GACCGTCCCGCCGCCCCAGAGTCTCTGTGTCTGTGGCCATCCGGGCTCCCCGCTCAC 7860
Qy 2621 SerGlnMetSerGlyValSerSerThrValGlyValValArgTyrSerCysIleGly 2640
Db 7861 TCCCAAGATCTCTGGAGACAGTTATCTGTGGAGAGAGAGTGGGTAAAGCTGCACTCGGC 7920
Qy 2641 LysArgThrLeuValGlyValSerThrArgMetCysGlyLeuAspGlyValSerThrGly 2660
Db 7921 AAGCGTACTCTGTGGAGAAACAGACCCGCAATGTGTGGCTGGAGACATGGACTGGC 7980
Qy 2661 SerLeuProHisCysSerGlyThrSerValGlyValCysGlyAspProGlyValProAla 2680
Db 7981 TCCCTCCCTCAGCTGCTCAGAGAACAGCGTGGAGATTGCGGTGACCTGGGATCCGGCT 8040
Qy 2681 HisGlyIleArgLeuGlyValAspSerPheAspProGlyThrValMetArgPheSerCysGly 2700
Db 8041 CATGGCATCCGTTTGGGGGACAGCTTTGATCCAGGACATGATGCGCTTCAGCTGTGAA 8100
Qy 2701 AlaGlyHisValLeuArgGlySerSerGlnArgThrCysGlnAlaAsnGlySerTrpSer 2720
Db 8101 GGTGGCCAGCTGCTCCGGGGATCGTCAGAGCGCACTGTCAAGCCATGGCTGTGAGC 8160
Qy 2721 GlySerGlnProGluCysGlyValIleSerCysGlyValAsnProGlyThrProSerAsnAla 2740
Db 8161 GGCTCGCAGCTGAGTGTGAGATGATCTCTGTGGAGAACCTGGGACTCCAAAGTAAGGCC 8220
Qy 2741 ArgValValPheSerAspGlyLeuValPheSerSerIleValTyrGluCysArgGlu 2760
Db 8221 CGAGTTGTGTCAATGATGGCGCTGTGTTTCTCCAGCTCATTCGTCTATGATGTCGGGAA 8280
Qy 2761 GlyTyrTyrAlaThrGlyLeuLeuSerArgHisCysSerValAsnGlyThrTrpThrGly 2780
Db 8281 GGATATCTAGCCACAGAGCTGCTGACGCGTCACGCTGCAATGGTACTGACAGAGCG 8340
Qy 2781 SerAspProGluCysLeuValIleAsnGlyValAspProGlyIleProAlaAsnGlyLeu 2800
Db 8341 AGTACCCTGAGTCCCTCGTATTAATGTGTGACCTGGGATTCAGCCAAATGGCCTT 8400
Qy 2801 ArgLeuGlnValAsnArgPheArgTyrAsnValThrValThrTyrGlnCysValProGlyTyr 2820
Db 8401 CGGTGGGCAATGATCTTCAAGTACAAACAACTGTGACATATGAGTGTCCCTGCTAT 8460
Qy 2821 MetMetGlnSerHisArgValSerValLeuSerCysThrTyrAspArgThrTrpAsnGly 2840
Db 8461 ATGATGAGTCACTAGAGATGTCTGTCTGAGCTGCAACAGGACCGGACATGGAATGGA 8520
Qy 2841 ThrLysProValCysValAlaLeuMetCysValAspProProLeuIleProAsnGlyLys 2860
Db 8521 ACCAAGCCCTGCAAAAGCTCTCATGTGCAAGCACTCCGCTATCCCAATGGGAAG 8580
Qy 2861 ValValGlySerAspPheMetTrpGlySerSerValThrTyrAlaCysLeuGlnGlyTyr 2880
Db 8581 GTGTGGGATCTGACTTCACTGTGGGCTCAAGTGTACTTATGCTGCTGGAGGGGTAC 8640
Qy 2881 GlnLeuSerLeuProAlaValPheThrCysGlnGlyValAsnGlySerTrpThrGlyValLeu 2900
Db 8641 CAGCTCTCCCTGCGCGGTGTCACTGTAGAGGAAATGGGTCTTGAACCGGAAGAGCTG 8700
Qy 2901 ProGlnCysPheProValPheCysGlyAspProGlyValProSerArgGlyValArgGly 2920
Db 8701 CCTCAGTGTTCCTGTGTGTGTGGGGATCTGTGGTCCCTCCGAGGAGAGAGAG 8760
Qy 2921 AspArgGlyPheSerTyrArgSerSerValSerPheSerCysHisProProLeuValLeu 2940
Db 8761 GACCGAGGCTTCTCTCAAGATCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 8820
Qy 2941 ValGlySerProArgArgPheCysGlnSerAspGlyThrTrpSerGlyThrGlnProSer 2960
Db 8821 GTGGGCTCTCCACGACGAGTTTTCAGATGAGGACATGGAGTGGACCCAGCCAGC 8880
Qy 2961 CysIleAspProThrLeuThrThrCysAlaAspProGlyValProGlnPheGlyIleGln 2980
Db 8881 TGCATGATCTCCAGCTTGCACAGTGTGGAGACCTGTGTGTGCACAGTTTGGATACAG 8940
Qy 2981 AsnAsnSerGlnGlyTyrGlnValGlySerThrValLeuPheArgCysGlnLysGlyTyr 3000
Db 8941 AACAAATTCAGAGGCTACAGAGTGGAGAACACAGTCTTCTCTGTGTGTGTGTGTGT 9000
Qy 3001 LeuLeuGlnGlySerThrThrArgThrCysLeuProAsnLeuThrTrpSerGlyThrPro 3020
Db 9001 GTGCTTCAGGGCTCCACACAGAGACCTGCTCCCAAACTGACCTGAGTGGAAACCCA 9060
Qy 3021 ProAspCysValProHisCysArgGlnProGluThrProThrHisAlaAsnValGly 3040
Db 9061 CCGATGCTGTCTCCCAACACATGCAAGGAGCCAGAGAGCCAAAGCATGCCAGTGGG 9120
Qy 3041 AlaLeuAspLeuProSerMetGlyTyrThrLeuIleThrProAlaArgArgAlaSerPro 3060
Db 9121 GCCCTGATTTGCTCTCATGGCTACAGGCTCATTACTCTGCGCAGAGAGGCTTCTCC 9180
Qy 3061 SerArgValAlaProSerThrAlaProAlaArgArgMetAlaAlaGlyGlnAlaSerArg 3080
Db 9181 TCAAGGATGGCTCCAGACACGCACTTCAAGCGGATGAGGAGCTGGACAGGCAAGCCGC 9240
Qy 3081 ProSerAlaTrpArgSerGlyProValGlyAspProSerThrLeuProGlySerHisArg 3100
Db 9241 CCATCTGCTGGAGGTCGGGCGCCAGTGGGAGAACCATCAACACTGCCGGAGCCACCGC 9300
Qy 3101 SerProLysPro 3104
Db 9301 TCACCCAGGCT 9312
Db 9301 TCACCCAGGCT 9312
RESULT 5
ADH71141
ID ADH71141 standard; DNA; 9551 BP.
XX
AC ADH71141;
XX
DT 25-MAR-2004 (first entry)
XX
DE Human gene of the invention NOV4 SEQ ID NO:37.
XX
KW ds; gene; human; cytotaxic; immunomodulator; neuroprotective; nootropic;
KW anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;
KW obesity; diabetes; infectious disease; metabolic syndrome X;
KW dyelipidaemia.
OS Homo sapiens.
XX
PN MO2003102155-A2.
XX
PD 11-DEC-2003.
XX
PF 03-JUN-2003; 2003MO-US017430.
XX
PR 03-JUN-2002; 2002US-0385120P.
PR 04-JUN-2002; 2002US-0385784P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386047P.
PR 06-JUN-2002; 2002US-0386376P.
PR 06-JUN-2002; 2002US-0386453P.
PR 06-JUN-2002; 2002US-0386846P.
PR 06-JUN-2002; 2002US-0387016P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386816P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0386942P.
PR 07-JUN-2002; 2002US-0386971P.
PR 07-JUN-2002; 2002US-0387262P.
PR 08-JUN-2002; 2002US-0296960P.
```

|    |             |                 |
|----|-------------|-----------------|
| PR | 10-JUN-2002 | 2002US-0387400P |
| PR | 10-JUN-2002 | 2002US-0387535P |
| PR | 11-JUN-2002 | 2002US-0387610P |
| PR | 11-JUN-2002 | 2002US-0387625P |
| PR | 11-JUN-2002 | 2002US-0387634P |
| PR | 11-JUN-2002 | 2002US-0387668P |
| PR | 11-JUN-2002 | 2002US-0387696P |
| PR | 11-JUN-2002 | 2002US-0387702P |
| PR | 11-JUN-2002 | 2002US-0387836P |
| PR | 11-JUN-2002 | 2002US-0387859P |
| PR | 12-JUN-2002 | 2002US-0387933P |
| PR | 12-JUN-2002 | 2002US-0387934P |
| PR | 12-JUN-2002 | 2002US-0387960P |
| PR | 12-JUN-2002 | 2002US-0388022P |
| PR | 12-JUN-2002 | 2002US-0388096P |
| PR | 13-JUN-2002 | 2002US-0389123P |
| PR | 14-JUN-2002 | 2002US-0389118P |
| PR | 14-JUN-2002 | 2002US-0389120P |
| PR | 14-JUN-2002 | 2002US-0389146P |
| PR | 14-JUN-2002 | 2002US-0389146P |
| PR | 17-JUN-2002 | 2002US-0389729P |
| PR | 17-JUN-2002 | 2002US-0389742P |
| PR | 18-JUN-2002 | 2002US-0389884P |
| PR | 19-JUN-2002 | 2002US-0390066P |
| PR | 19-JUN-2002 | 2002US-0390209P |
| PR | 21-JUN-2002 | 2002US-0390763P |
| PR | 17-JUL-2002 | 2002US-0396706P |
| PR | 06-AUG-2002 | 2002US-0401628P |
| PR | 09-AUG-2002 | 2002US-0402156P |
| PR | 09-AUG-2002 | 2002US-0402256P |
| PR | 09-AUG-2002 | 2002US-0402389P |
| PR | 12-AUG-2002 | 2002US-0402786P |
| PR | 12-AUG-2002 | 2002US-0402816P |
| PR | 12-AUG-2002 | 2002US-0402821P |
| PR | 12-AUG-2002 | 2002US-0402832P |
| PR | 13-AUG-2002 | 2002US-0403448P |
| PR | 13-AUG-2002 | 2002US-0403459P |
| PR | 13-AUG-2002 | 2002US-0403531P |
| PR | 13-AUG-2002 | 2002US-0403532P |
| PR | 13-AUG-2002 | 2002US-0403563P |
| PR | 13-AUG-2002 | 2002US-0406317P |
| PR | 15-AUG-2002 | 2002US-0403617P |
| PR | 26-AUG-2002 | 2002US-0406182P |
| PR | 26-AUG-2002 | 2002US-0406355P |
| PR | 27-AUG-2002 | 2002US-0406240P |
| PR | 12-SEP-2002 | 2002US-0410084P |
| PR | 20-SEP-2002 | 2002US-0412528P |
| PR | 23-SEP-2002 | 2002US-0412731P |
| PR | 30-SEP-2002 | 2002US-0414801P |
| PR | 30-SEP-2002 | 2002US-0414839P |
| PR | 30-SEP-2002 | 2002US-0414840P |
| PR | 30-SEP-2002 | 2002US-0414954P |
| PR | 09-OCT-2002 | 2002US-0417186P |
| PR | 09-OCT-2002 | 2002US-0417406P |
| PR | 23-OCT-2002 | 2002US-0420639P |
| PR | 28-OCT-2002 | 2002US-0421156P |
| PR | 31-OCT-2002 | 2002US-0422690P |
| PR | 01-NOV-2002 | 2002US-0423130P |
| PR | 05-NOV-2002 | 2002US-0423798P |
| PR | 05-NOV-2002 | 2002US-0423798P |
| PR | 12-NOV-2002 | 2002US-0425453P |

[illegible]

|  |   |
|--|---|
| XX   | WP1; 2004-081935/08.  |
| DR   | P-PsDB; ADH71142.   |
| XX   | New NOXV polypeptides and nucleic acid molecules useful for preventing or |
| PT   | treating NOXV-associated disorders, e.g. cancer, diabetes, infection or   |
| PT   | obesity, and in chromosome mapping, tissue typing or pharmacogenomics.    |
| XX   |   |
| PS   | Example 4; SEQ ID NO 37; 1880bp; English.                                 |
| CC   | The invention relates to a novel isolated polypeptide (NOXV). A           |
| CC   | polypeptide of the invention has cyrostatic, immunomodulator,             |
| CC   | neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and   |
| CC   | antihypaemic activity, and may have a use in gene therapy, and as a       |
| CC   | vaccine. The polypeptides are encoded by NOXV polynucleotides comprising  |
| CC   | any of the 303 fully defined nucleotide sequences given in the            |
| CC   | specification. The polypeptide is useful in the manufacture of a          |
| CC   | medicament for treating a syndrome associated with a human disease. The   |
| CC   | polypeptide, polynucleotide and antibody are useful in diagnosing,        |
| CC   | treating or preventing NOXV-associated disorders, e.g. cancer, cachexia,  |
| CC   | Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious   |
| CC   | diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are   |
| CC   | further used as hybridisation probes, in chromosome mapping, tissue       |
| CC   | typing, preventive medicine, and pharmacogenomics. The present sequence   |
| CC   | encodes a NOXV polypeptide of the invention.                              |
| XX   |   |
| SEQ  | Sequence 9951 BP; 2118 A; 2982 C; 2697 G; 2154 T; 0 U; 0 Other;           |
| Alignment Scores:                            |   |
| Pred. No.:                                   | 0 Length: 9951  |
| Score:                                       | 16512.00 Matches: 3037  |
| Percent Similarity:                          | 97.78% Conservative: 1  |
| Best Local Similarity:                       | 97.75% Mismatches: 2  |
| Query Match:                                 | 97.22% Indels: 69   |
| Dbs:   | 12 Gaps: 3  |
| US-10-016-248-2 (1-3104) x ADH71141 (1-9951) |   |
| OY   | 1 MetLaGIAlAProPProAlaLeuLeuPProCysSerLeuIleSerAspCysCys 20               |
| Db   | 1 ATGGGCGGCGCCCCCTCCCCCGCCTTGCTGCTGCTTGCAGTTGAATCTCAGACTGCTGT 60          |
| OY   | 21 AlaSerAngInaIGHisSerValIGlyValIGlyProSerGluLeuValILysGLNile 40         |
| Db   | 61 GCTAGCATACGACGACACTCCGTGGGGGTAGGACCCTCCGAGCTAGTCAGAACAATT 120          |
| OY   | 41 GluLeuLysSerArgLYalValLysLeuMetProSerLysAspAsnSerGlnLysThrSer 60       |
| Db   | 121 GAGTTGAAGTCTCAGAGGTGTAAACCTGATGCCCGCAAGAACACACCCAGAGACCTCT 180        |
| OY   | 61 ValLeuThrGlnValIGlyValSerGlnGlyHisAsnMetCysProAspProGlyIlePro 80       |
| Db   | 181 GTGTTAACTCGGTTGGTGTGTCCCAAGAAATATATGTGTCCAGACCCCTGGGATACCC 240        |
| OY   | 81 GluArgGlyLysArgLeuGlySerAspPheArgLeuGlySerSerValGlnPheThrCys 100       |
| Db   | 241 GAAGGGGCAAAAGACATAGGCTCCGATTTACAGTTAGATTCACCGCTCCAGTTACCTGC 300       |
| OY   | 101 AsnGlnGlyTYrAspLeuGlnGlySerLysArgIleThrCysWetLysValSerAspMet 120      |
| Db   | 301 AACGAGGGCTATGACCTGCAAGGGGTCCAAAGCGGATCACCTGTATGAAGATGACGCAATG 360     |
| OY   | 121 PheAlaAlaITPserAspHisArgProValCYarArgAlaArgMetCysAspAlaHisLeu 140     |
| Db   | 361 TTTCGGCGCTTGACAGCAACACAGGCCAGTCTGCCAGACCCGCAAGTGTGATGCCACCTT 420      |
| OY   | 141 ArgGlyProSerGlyIleIleThrSerProAsnPheProIleGlnTYrAspAsnAla 160         |
| Db   | 421 CGAGGCCCCCTGGGACATATACCTCCCACATTTCCCATTCAGTATGACACAAATCCA 480         |
| OY   | 161 HisCysValITrpIleIleThrAlaLeuAsnProSerLysValIleLysLeuAlaPheGlu 180     |
| Db   | 481 CACTGTGTGTGGATCATACAGGACTCAACCCCTCCAAAGGATCAAGAGCTGCTTGTAG 540        |

QY 181 GluPheAspLeuGluValArgGlyTyrAspThrLeuThrValGlyAspGlyGlyGlnAspGly 200  
DB 541 GAGTTGATTTGGAGAGGGGCTATGACCTGACCGGTGGTGAATGGTGTGATGAGATGAGG 600  
QY 201 AspGlnLysThrValLeuTyrMetSerGlnAsnAlaCysSerAspSerProHisThrPro 220  
DB 601 GACCAAGACAGATTTCTTACATGTTCTAAATGCTTCAGATGACAGCCCTTCACACCCCA 660  
QY 221 GlySerArgIleIleProGluSerMetSerGlyAspIleTyrArgGlnLysTyrThrValLeu 240  
DB 661 GGCTCTGCGATCCCAAGAGACATGCTGGGGACATCTGAGGCAAGAAATGACCTGATCTT 720  
QY 241 GlnIleCysArgAspIleSerSerSerAspAlaArgSerGlySerValArgLysSerPro 260  
DB 721 GAGATCTGTCGATGACATTTAGCATTCAGATGCAAGGTCCAGGTTCAGTAGAAGCTTCCA 780  
QY 261 LysThrSerAsnAlaValGlnLeuValAlaProGlyThrGlnLysGlnGlySerCys 280  
DB 781 AAGACTTCTTATGCTGTGGAACTTGTGCTCTGGGACAGAGATGCAAGGCAAGTTGC 840  
QY 281 GlyAspProGlyIleProAlaTyrGlyValArgArgGlyLysArgPheHisGlyAsp 300  
DB 841 GGTGACCTGGCATCTGTCATATGGCCGAGAGGAGGCTCCGGTTTCACACAGGTGAC 900  
QY 301 ThrLeuLysPheGlnCysGlnProAlaPheGlnLeuValGlyGlnLysAlaIleThrCys 320  
DB 901 ACACCTCAAGTTGAGTGCACAGCCGCTTTAGGCTGGTGGGACAGAAAGCATACATCC 960  
QY 321 GlnLysAsnAsnGlnTyrPheSerAlaLysLysProGlyCysValPheSerCysPheAsn 340  
DB 961 CAAGAATTAACCATGATGGTGGCTTAAGAAAGCAGAGCTGGTGTCTCTGCTCTTCAAC 1020  
QY 341 PheThrSerProSerGlyValValLeuSerProAsnTyrProGluAspTyrGlyLysAsnHis 360  
DB 1021 TTCACAGCCCGCTGTGGGTTGTCTGTCTCCCACTACCAAGAGCATAGGCAACAC 1080  
QY 361 LeuHisCysValIleTyrLeuIleLeuAlaArgProGluSerArgIleHisLeuAlaPheAsn 380  
DB 1081 CTCACAGTGTGCTGCTGCTCATCTGCGCAGGCTGAGAGCCGATCCACTGAGCTTCAAC 1140  
QY 381 AspIleAspValGluProGlnPheAspPheLeuValIleLysAspGlyValAlaThrAlaGlu 400  
DB 1141 GACATTAACGTCGAGGCTCAGTTGATTTCTGTGTCATCAAGATGGGGCAGCCGCGAG 1200  
QY 401 AlaProValLeuGlyThrPheSerGlyAsnGlnLeuProSerSerIleThrSerSerGly 420  
DB 1201 GCGCCGCTGCTGGGCACTTCTCAGGAAACAGCTTCCCTCCATCCACAGCAGTGGC 1260  
QY 421 HisValAlaArgLeuGluPheGlnThrAspHisSerThrGlyLysArgGlyPheAsnIle 440  
DB 1261 CACGTGCCCCGTCTCGAGTTCAGACTGACACTCCACAGGAAAGGGGCTTCAACATCC 1320  
QY 441 ThrPheThrThrPheArgHisAsnGluCysProAspProGlyValProValAsnGlyLys 460  
DB 1321 ACTTTTACCACTTCCGACAAACAGAGTCCCGGATCCGCGCTTCAGTAATGGCAAA 1380  
QY 461 ArgPheGlyLysAspSerLeuGlnLeuGlySerSerIleSerPheLeuCysAspGlnGlyPhe 480  
DB 1381 CGGTTTGGGACAGCTCCAGCTGGGACACTGATCTCCCTCTGTGATGAAGGCTTC 1440  
QY 481 LeuGlyThrGlnLysSerGluThrIleThrCysValLeuLeuGlnGlySerValAlaTyr 500  
DB 1441 CTGGGACCTCAGAGGCTCAGAGCACTGCTGCTTCCTTAAGAGAGGCGAGCTGTCTGG 1500  
QY 501 AsnSerAlaValLeuArgCysGlnAlaProCysGlyGlyHisLeuThrSerProSerGly 520  
DB 1501 AAGAGGCTGTGTGGGTGAGGCTCCCTGTGGGTGACCTGATCCGCGCCAGCGGC 1560  
QY 521 ThrIleLeuSerProGlyTyrProGlyPheTyrLysAspAlaLeuSerCysAlaTyrVal 540  
DB 1561 ACCATCTCTCTCCGGGCTGGCTGCTTCTTCAAGATGCTCTGAGCTGTGGCTG 1620

QY 541 IleGlnAlaGlnProGlyTyrProIleLysIleThrPheAspArgPheLysThrGlnVal 560  
DB 1621 ATTGAGGCCACAGCAGGCTACCCCATCAAAATCACTTGACAGATTTCAAAACCGAGGTC 1680  
QY 561 AsnTyrAspThrLeuGlnValArgAspGlyArgThrTyrSerAlaProLeuIleGlyVal 580  
DB 1681 AACTATGACACCTCGAAGATGACGATGGGCGGACTTACAGCCGCTTGATCCGGGTT 1740  
QY 581 TyrHisGlyThrGlnAlaProGlnPheLeuIleSerThrSerAsnTyrLeuLeu 600  
DB 1741 TACCAAGGACCCAGGATTCACAGTTCCTCATCAGACAGCAACATCACTTACCTCTCC 1800  
QY 601 PheSerThrAspLysSerHisSerAspIleGlyPheGlnLeuArgTyrGlnThrIleThr 620  
DB 1801 TTCTTACCGACAAAGATCACTCGGACATCGGCTTCCAGCTCCGTTGAGACTTAA 1860  
QY 621 LeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGlnArgHisGlyAsn 640  
DB 1861 CTGCACTCAGACACCTGCTGTGATCCAGAAATCCAGTAAATGACAGCGTCATGGAAAT 1920  
QY 641 AspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrThrLeuSerAsp 660  
DB 1921 GACTTCTAGTGGGCGCTGTGATCTTCACTGTGACTCGGGCTTACACATTAAGTAC 1980  
QY 661 GlyGluProLeuGlnCysGluProAsnPheGlnTyrPheSerArgAlaLeuProSerCysGlu 680  
DB 1981 GGGAGCCCTCGAGATGAGGCCAACTTCCAGTGAAGCCGGGCTCCGACAGTTGTGA 2040  
QY 681 AlaLeuCysGlyGlyPheIleGlnLysSerSerGlyThrIleLeuSerProGlyPhePro 700  
DB 2041 GCTCTGTGTGGCTTATTCAGAGCTTCAGAGGCAACATCTGTGGCCAGGGTCTCCCT 2100  
QY 701 AspPheTyrProAsnAsnLeuAsnCysThrTyrIleIleGlnThrSerHisGlyLysGly 720  
DB 2101 GACTTCTACCCCAACATTAATGAATGACCTGATTTCAAAATCATCTGAGGAAGGT 2160  
QY 721 ValPhePheThrPheHisThrPheHisLeuGlnSerGlyHisAspTyrLeuLeuIleThr 740  
DB 2161 GTGTTCTTCACTTTCACACCTTCCAGCTGGAAGGTGGCAATGACTCTCCATCATCACT 2220  
QY 741 GluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760  
DB 2221 GAGAAAGGCACTTCCACCAAGCCCTGAGGCAAGTAACTGATTCGGCTGCCAGCTCCC 2280  
QY 761 IleSerIleGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780  
DB 2281 ATCAGCGCTGGCTTATGGCACTTCACTGCGCAGGCTCGCTTCAATCTGATTTCTCC 2340  
QY 781 MetSerTyrGlnGlyPheAsnIleThrPheSerGluTyrAspLeuGluProCysGlnGlu 800  
DB 2341 ATGTCATTAAGAGATTCACACATCACTTTCAGAGTACGACTTGGAGCCCTTGAAGAG 2400  
QY 801 ProGluValProAlaTyrSerIleArgLysGlyLeuGlnPheGlyValAlaAspThrLeu 820  
DB 2401 CCCAGGCTCCACCTTACAGCATCCGGAAGGGCTTCAAGTTTGGCGTGGCGCACACTTG 2460  
QY 821 ThrPheSerCysPheProGlyTyrArgLeuGlnGlyThrAlaArgIleThrCysLeuGly 840  
DB 2461 ACCTTCTCTGCTTCCCGGGATACCGTCTGGAAGGAGCAGCGCGCATCACTGCCGGGG 2520  
QY 841 GlyArgArgArgLeuTyrPheSerProLeuProArgCysValAlaGlnCysGlyAsnSer 860  
DB 2521 GGCAGAGCGGCGCTGTGAGCTGCTGTGCCAAGGTGTGTGTCTGAGTGTGGAAATTA 2580  
QY 861 ValThrGlyThrGlnLysThrLeuLeuSerProAsnPheProValAsnTyrAsnAsn 880  
DB 2581 GTACAGGACCTCAGGGTACTTGTGCTGCCCACTTCTCTGTGAACATAATTAACAT 2640  
QY 881 HisGluCysIleTyrSerIleGlnThrGlnProGlyLysGlyIleGlnLeuLysValArg 900  
DB 2641 CATGATGATCTACTCATCCAGACCCAGCGAGGAAGGGAATTCAGCTGAAGCCAGG 2700  
QY 901 AlaPheGlnLeuSerGlnGlyAspValLeuLysValTyrAspGlyAsnAsnAsnSerAla 920

```
Db 2701 GCATTCGAACCTCCGAGAGAGATGTCTCAAGGTTTATGATGGCAACAACATCCGCC 2760
Qy ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValIThrLeuAsnSerThrSer 940
Db 2761 CGTTTGCTGGAGATTTTAAAGCATTCCTGAGATGATGGGGTGACTTTTGAACAGACATCC 2820
Qy SerSerLeuTrpLeuAspPheIleThrAspAlaGluAsnThrSerLysGlyPheGluLeu 960
Db 2821 AGCAGCTGTGGCTTGAATTTATCACTGATCTGAAAACACCAAGAGGGCTTTGAACCTG 2880
Qy HisPheSerSerPheGluLeuIleLysCysGluAspProGlyIThrProLysPheGlyThr 980
Db 2881 CACTTTTCCAGCTTTGAACCTCATCAATGATGAGAGCCAGAGACCCCAAGATTGGCTAC 2940
Qy LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCysAspProGly 1000
Db 2941 AAGGTTCAATGATGAGAGTCATTTTGCAGAGAGCTCCGTCTTCAGCTGTGACCTGGA 3000
Qy TyrSerLeuAspGlySerGlyLysGlyLysLeuLeuCysLeuSerSerGlyLysArgThrTrpAsp 1020
Db 3001 TACAGCTCTGGGGGTAGTGAAGAGCTGTGTCTGAGTGAAGAGCCGGACCTGGAGAC 3060
Qy ArgProLeuProThrCysValAlaGlyCysGlyGlyIThrValArgGlyGlyValSerGly 1040
Db 3061 CGGCTCTGGCCCACTGTGTGCGGAGTGTGAGAGGACAGTGAAGAGAGAGTGTCCGGGG 3120
Qy GluValLeuSerProGlyIThrProAlaProGlyGluHisAsnLeuAsnCysIleTrpThr 1060
Db 3121 CAGGTGTCTGACCCGGGTATCCAGCTCCCTATGACACAACTTCACTGATCTGAGACC 3180
Qy IleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGluGlu 1080
Db 3181 ATCAAGACGAGGCGCGGCTGCAACATTTGGCTTCACTTCTGTGTGTGACACAGAGAG 3240
Qy ValHisAspValLeuArgIleTrpAspGlyProValGlySerGlyValLeuLeuLysGlu 1100
Db 3241 GTTCACAGACGTGTGCGGATCTGGGATGGGCTGTGAGAGAGCGGGGTTCTGTAAAGAG 3300
Qy LeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValIleLeuGln 1120
Db 3301 CTGATGGGCGCGGCTGCGCCAGAGACCTGATAGCACTTCACTGCTGTGTCTCTCAG 3360
Qy PheSerThrAspPhePheThrSerLysGlnGlyPheAlaIleGlnPheSerValSerThr 1140
Db 3361 TTTCGACTGACTTCTTTCACAGACAGAGAGGCTTTGCAATTCATTTTCCAGTCCACA 3420
Qy AlaThrSerCysAsnAspProGlyIleProGluHisGlnGlySerArgSerGlyAspSerTrp 1160
Db 3421 GCACGCTCTGCAATGACCTCGGGATCCCGCAAGATGGGAGTCCGAGTGTGACAGATTGG 3480
Qy GluAlaGlyAspSerThrValPheGlnCysAspProGlyIThrAlaLeuGlnGlySerAla 1180
Db 3481 GAACCGCGGCACTCCACAGTGTTCAGTGTGACCTGCTGACGCTGACAGGAGATGCA 3540
Qy GluIleSerCysValLysIleGluAsnAspPhePheTrpGlnProSerProProThrCys 1200
Db 3541 GAGATCAGCTGTGAGAGATGAGAAACAGTTCCTTCTGGCAGCCAGCCGCAACAATGC 3600
Qy IleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTrp 1220
Db 3601 ATGCTCTCTGCGGGGAGAGACTGACAGGACCATCTGAGATCACTCTCCCAAAATTAC 3660
Qy ProGluProIleProProGlyLysGlyCysAspTrpLysValIThrValSerProAspTrp 1240
Db 3661 CCAAAACCTTACCCGCGCAGAGAGAGTGTGACAGAAAGTGAACCGTCTCCACAGATAC 3720
Qy ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyIThrAspPheLeuHisIle 1260
Db 3721 GTCAATGCGCTGTGATTTAACTTTAACTTGAAGGCTGCGTAAAGACTTCTCCATATTC 3780
Qy TyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheIleGlySerGlnLeuPro 1280
Db 3781 TACAGCAGAGCGAGACTCTCAGCCCTCATAGGAAGCTTATGGCTCCAGCTCCCA 3840
Qy GlyValGlyIleGlySerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal 1300
Db 3841 GGCGCAATTAAAGACAGACAAACAGCTCTTCTTCCCTTCCGACAGATGATCTGTG 3900
Qy SerAsnAlaGlyPheValIleAspIleThrGluAsnProArgGlySerCysPheAspPro 1320
Db 3901 AGCAATGCTGGCTTCGTCATTTGACTATACAGAAAACCCGGGAGATGATTTTTGATCT 3960
Qy GlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeuGlySerSerValThr 1340
Db 3961 GGTTCCATCAAGAACGACACACGGGTGGGTCCGACCTGAAGCTGGGCTCTCCGTACCC 4020
Qy TyrTrpCysHisGlyGlyIThrGluValGluGlyIThrSerThrLeuSerCysIleLeuGly 1360
Db 4021 TACTACTGCCACGGGGGCTACGAAGTGAAGGACCTCGACCTGACCTCATCTCTGGG 4080
Qy ProAspGlyLysProValTrpAsnAspProArgProValCysThrAlaProCysGlyGly 1380
Db 4081 CTGATGGAGAACCCGCTGTGGAACAATCCCGGCACTGTGACAGCCCTGTGGGGGA 4140
Qy GlnIleValGlySerAspGlyValIleLeuSerProAsnTrpProGluAsnTrpThrSer 1400
Db 4141 CAGTATGGGTTTCGACGAGATGTCTTGTCCCCCACTAACCCCGAAGTACACAGAT 4200
Qy GlyGlnIleCysLeuTrpPheValIThrValProLysAspTrpValValPheGlyGlnPhe 1420
Db 4201 GGAACAGATCTGCTGTGATTTTGTTCATGTGCCAACAGGACATGATGTGTGGCAGATT 4260
Qy AlaPhePheIleThrAlaLeuAsnAspValAlaGlyValHisAspGlyHisSerGlnHis 1440
Db 4261 GCCTCTTTACACAGCGCTTCACAGAGCTGTGAGTTTACACAGCGCCACAGCAGACAC 4320
Qy SerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGlyLeuSerLeuProLeuAlaThr 1460
Db 4321 TCGGGGCTCTCAGCTCCCTCGGGGCTCCCATACAGAGAAATCACTGCCCTTGGCCACC 4380
Qy SerAsnGlnValLeuIleLysPheSerAlaLysGlyLeuAlaProAlaArgGlyPheHis 1480
Db 4381 TCCAAATCAATTTCTCATTAAGTTCAAGGCCCAAGAGGCTTCACACAGCAGAGGCTTCC 4440
Qy PheValTrpGlnAlaValProArgThrSerAlaThrGlnCysSerSerValProGluPro 1500
Db 4441 TTTGTCTACCAAGCGTTCCTTCAACAGGCGCCACAGAGTGCAGCTGTGTGCGGAACCC 4500
Qy ArgTrpGlyLysArgLeuGlySerAspPheSerValGlyValIleValArgPheGluCys 1520
Db 4501 CGTATGGAAGAGGCTGGGCAATGACTTCTCGGTGGGGGCAATCGTCCGCTTCGAATGC 4560
Qy AsnSerGlyIThrAlaLeuGlnGlySerProGluIleGlyCysLeuProValProGlyAla 1540
Db 4561 AACTCCGCTATGCTCCGCGAGGGGTCCGCAAGAAATCAGAGCTCTCCCTGTGCTGGGGCC 4620
Qy LeuAlaGlnTrpAsnValSerAlaProThrCysValValProCysGlyGlyAsnLeuThr 1560
Db 4621 TTGGCCCAATGAGATGTTCTCAAGGCGCCACGATGTGTGTCGCTGTGAGGCAACTTACA 4680
Qy GluArgArgIleThrIleLeuSerProGlyPheProGluProTrpTrpLeuAsnSerLeuAsn 1580
Db 4681 GAGGCGAGGGGCAACCATCTGTCCCTCGCTTCCAGAGCGATCACTCAACAGCTTCAAC 4740
Qy CysValTrpLysIleValValProGluGlyValGlyIleGlnIleGlnValValSerPhe 1600
Db 4741 TGTGTGGAAGATCTGGGTCCCGGAAGGGCTGGGATTCAGATTCCAAGTTGTCAAGTTT 4800
Qy ValThrGluGlnAsnTrpAspSerLeuGluValPheAspGlyValHisAspAsnThrValThr 1620
Db 4801 GTGACAGAGAGAACTGGAGCTGCTGGAAGTATTTGATGTGTGACATTAACCTGTAAAC 4860
Qy MetLeuGlySerPheSerGlyIThrThrValProAlaLeuLeuAsnSerThrSerAsnGln 1640
Db 4861 ATGCTGGGAGATTTCTCAGAGAACACCGTGTGCTGTCTGAACAGCACTTCAACAG 4920
```

|    |      |  |      |
|----|------|--|------|
| Qy | 1641 | LeuTyrLeuHisPheTyrSerAspIleSerValSerAlaGlyPheHisIleuGluTyr         | 1660 |
| Db | 4921 | CTTACCTTCATTCTTACTACAGATATCAGGATCTGACGGCTTCCACTTGGAGTAC            | 4980 |
| Qy | 1661 | LysThrValGlyLeuSerSerCysProGluProAlaValProSerAsnGlyValIleThr       | 1680 |
| Db | 4981 | AAATCAGGTGGGCTGAGCAGTTGTCCGGAACCTGGTGTCCAGTACCGGGTGAAGCT           | 5040 |
| Qy | 1681 | GlyGluArgTyrLeuValAsnAspValIleSerPheGlnCysGluProGlyTyrAlaLeu       | 1700 |
| Db | 5041 | GGGAGCCCTACTTGTGATGATGTGTCTTCCAGTGTGACCCGGATATGCTCTC               | 5100 |
| Qy | 1701 | GlnGlyValAlaHisIleSerCysMetProGlyTyrThrValArgAspTyrAspTyrPro       | 1720 |
| Db | 5101 | CAGGGCCAGCCCACTCTCCGATGCCCCGGAACAGTGGCAGTGAATCACTCTCTC             | 5160 |
| Qy | 1721 | ProLeuSerValAlaGlnCysGlyGlyThrValGluGluMetGluGlyValIleLeuSer       | 1740 |
| Db | 5161 | CCACTCTGTATTTGACAGTGTGGGGAAACAGTGAGAGATGAGAGGGGTGATCTGAGC          | 5220 |
| Qy | 1741 | ProGlyPheProGlyAsnTyrProSerAsnMetAspCysSerTyrIleAlaLeuPro          | 1760 |
| Db | 5221 | CCGGCTTCCAGGAACTACCCCACTGATGACAGTGTCTGAAAAATAGCACTGCC              | 5280 |
| Qy | 1761 | ValGlyPheGlyValAlaHisIleGlnPheLeuAsnPheSerThrGluProAsnHisAspTyr    | 1780 |
| Db | 5281 | GTGGGCTTTGGAGCTCAGATCAGTCTCTGAATCTTCCAGCCGACCAACCACTAC             | 5340 |
| Qy | 1781 | IleGluIleArgAsnGlyProTyrGluThrSerThrMetMetGlyArgPheSerGlySer       | 1800 |
| Db | 5341 | ATAGAAATCCGGAATGAGCCCTTATGAGCCAGCCGATGAGGAAATTCAGTGGAAAGC          | 5400 |
| Qy | 1801 | GluLeuProSerSerLeuLeuSerThrSerHisGluThrThrValTyrPheHisSerAsp       | 1820 |
| Db | 5401 | GAGCTTCCAAAGCTCCCTCTCTCCACGTCACAGAACCAACCGTATTTCCACAGGAC           | 5460 |
| Qy | 1821 | HisSerGlnAsnArgProGlyPheGlyLeuGluTyrGlnAlaTyrGluLeuGlnGlyCys       | 1840 |
| Db | 5461 | CATCTCCAGATCGGACCAAGATTCAGCTGAGATGAGGCTATGAACTTCAAGAGTGC           | 5520 |
| Qy | 1841 | ProAspProGluProPheAlaAsnGlyIleValArgGlyValGlyTyrAsnValGlyGln       | 1860 |
| Db | 5521 | CCAGATGGACCAACCGGAATCGGACATGTGAGGGAGCTGCTACCAACGCGGACAA            | 5580 |
| Qy | 1861 | SerValThrPheGluCysLeuProGlyTyrGlnLeuThrGlyHisProValLeuThrCys       | 1880 |
| Db | 5581 | TCACTGACCTTTCAGTGTCTCCCGGGGTATCAATTGACTGGCCACTGCTCACTGT            | 5640 |
| Qy | 1881 | GlnHisGlyThrAsnArgAsnTyrAspHisProLeuProGlyGlyValProCysGly          | 1900 |
| Db | 5641 | CAACATGGACCAACCGGAATCGGACATCGGACCAAGGTGAAATCTCTTGTGTGC             | 5700 |
| Qy | 1901 | GlyAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyrSer       | 1920 |
| Db | 5701 | GGGAACATCACTTCTTCCAGGCACTGTGTACTCCCGGGGTCTCCCTGACCCGTACTCC         | 5760 |
| Qy | 1921 | SerSerGlyAspCysValTyrPheIleThrValProIleGlyHisGlyValArgLeuAsn       | 1940 |
| Db | 5761 | AGCTCCCAAGACGTGTCTGGCTGATCACCCTGCTCCATTTGGCCATGGCTCCGCTCAAC        | 5820 |
| Qy | 1941 | LeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTyrAspGlyProGln       | 1960 |
| Db | 5821 | CTCAGCTCTCTGACAGACAGGCTCTGAGATTTCAATCAATCAATCTGGAGTGGACACAG        | 5880 |
| Qy | 1961 | GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaValLeuThrValGlnSer       | 1980 |
| Db | 5881 | CAAAACGACCAACGGCTCGGCTGCTTTCACACCCGACACATGGCCAAAGAAAACGTGACAGT     | 5940 |
| Qy | 1981 | SerSerAsnGlnValLeuLeuLysPheHisArgAspAlaIleThrGlyGlyIlePheAla       | 2000 |
| Db | 5941 | TCATCCAAACCAAGTCTCTGCTCAAGTTCCACGTGATGACCAAGGGGGGATCTTGCCTC        | 6000 |
| Qy | 2001 | IleAlaPheSerAlaTyrProLeuThrLysCysProProThrThrIleLeuProAsnAla       | 2020 |
| Db | 6001 | ATAGCTTCTCGGCTTATTCACATCACCACCAATGCCCTCTCCACCACTCTCCCAACGCTC       | 6060 |
| Qy | 2021 | GluValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrArgCysLeu       | 2040 |
| Db | 6061 | GAAGTGTGTCAACAGATGATGAGAAATTCATATATGAGTGCATGTGATGATGATGCTC         | 6120 |
| Qy | 2041 | ProGlyPheThrLeuValGlyAsnGluIleLeuThrCysValLeuGlyTyrThrLeuGln       | 2060 |
| Db | 6121 | CTGGCTTATACCTTATGTGGGAAATGAAATTTCTGACCTGCAAACTTGGAACTTACCTGCAG     | 6180 |
| Qy | 2061 | PheGluGlyProProProIleCysGluValHisCysProThrAsnGluLeuLeuThrAsp       | 2080 |
| Db | 6181 | TTTGAAGACCAACCCCGATATGTGAGTGCATGTCCCAAAAGAGCTTGTGACAGAC            | 6240 |
| Qy | 2081 | SerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCys       | 2100 |
| Db | 6241 | TTCACAGGCGGTATCTGTGAGCCAGCTAACCTTGAAAGCTATCCCACTTCCAGACTGC         | 6300 |
| Qy | 2101 | SerTyrPheValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeu       | 2120 |
| Db | 6301 | TCTTGGCTGTGAGAGTGGAGCCGACCTATACATCTCTCCACAGTGAAGTACTCTCTC          | 6360 |
| Qy | 2121 | SerGluGlyGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeu       | 2140 |
| Db | 6361 | ACGAGAAACATATGATGAGATTTGAGATTTTGTGATGCTCATGACAGACAGATCTCTG         | 6420 |
| Qy | 2141 | LeuValAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerSerAsnSer       | 2160 |
| Db | 6421 | CTGAAGGCTCAGTGGGAATTACTCACTCTCCCTGATTTGACACAGCTCAAGCAACTCT         | 6480 |
| Qy | 2161 | ValTyrLeuAlaGlyTyrSerSerAspHisAlaTyrAsnArgGlyPheLysIleArgTyr       | 2180 |
| Db | 6481 | GTGATCTGCGTGTGATCATGTGATCAGCTCAACATCGGAAGGCTTCAAGATCGCTAT          | 6540 |
| Qy | 2181 | SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr       | 2200 |
| Db | 6541 | TCA-----   | 6543 |
| Qy | 2201 | SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly       | 2220 |
| Db | 6543 | -----  | 6543 |
| Qy | 2221 | HisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisIleuThrSerGluAlaIle      | 2240 |
| Db | 6543 | -----  | 6543 |
| Qy | 2241 | ProLeuGlyGlnAlaLeuSerCysGlyLeuProGluAlaProLysAsnGlyMetValPhe       | 2260 |
| Db | 6544 | -----GCTCTTCTGTGGCTTCTGAGAGCCCTCCAGAAATGGAAATGTGTCTT               | 6591 |
| Qy | 2261 | GlyLysGluTyrThrValGlyThrLysAlaValTyrSerCysSerGluGlyTyrHisIleu      | 2280 |
| Db | 6592 | GGCAGGAGATGACAGATGGGAAACCAAGGCCATGTACAGCTGAGTAAGGCTACCACTCC        | 6651 |
| Qy | 2281 | GlnAlaGlyValAlaGlnAlaThrAlaGluCysLeuAspThrGlyLeuThrPheSerAsnArgAsn | 2300 |
| Db | 6652 | CAGGACAGGCTTATGAGCTGACAGATGTCTGGAACACAGGCTTATGAGCAACCGCAAT         | 6711 |
| Qy | 2301 | ValProProGlnCysValProValThrCysProAspValSerSerIleSerValGluHis       | 2320 |
| Db | 6712 | GTCCCAACCAAGTGTCTGTGACTTGTCTGATGTCAAGTACATGATGATGATGATGATGAT       | 6771 |
| Qy | 2321 | GlyArgTyrArgLeuIlePheGluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIle       | 2340 |
| Db | 6772 | GGCCGATGAGGCTTATCTTTAAGACACAGTACAGTTCACAGGCTTATGAGTGTCTCATC        | 6831 |
| Qy | 2341 | CysAspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLys       | 2360 |
| Db | 6832 | TGTGACCTGTGGCTACTACTGTGCAAGAGGTATCCGCTGTCAAGGCCAATGCGCAAA          | 6891 |
| Qy | 2361 | TyrSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeuProIle       | 2380 |

Db 6892 TGGAGCCCTGGGGAGCTACGCCACCTGCCGAATCATCTCTGGAGAGCTCCGATT 6951  
Qy 2381 ProProAenGIYHIAHxIGIYThrLeuSerValTYRGIYAlaThrAlaIlePheSer 2400  
Db 6952 CCCCCAATGGCCACCGGATCGGAACACTGTCTGTACGGGGCAACGCCATCTTCTCC 7011  
Qy 2401 CybAenSerGIYTYrThLeuValGIYSerArgValArgIuCyMetAlaAenGIYLeu 2420  
Db 7012 TGCATTCGGATACACACTGGTGGGCTCCAGGGGCTGTAGTGCATGGCCAAATGGGCTC 7071  
Qy 2421 TrpSerGIYSerGIYValArgCyLeuAlaGIYHISCySGIYThrProGIYProIleVal 2440  
Db 7072 TGGATGGCTCTGAAGTCGGCTGCTTGTCTGGAGACCTGTGGGAGCTCCGAGCCATTGTC 7131  
Qy 2441 AenGIYHISIleAenGIYLeuAenTYrSerTYrArgGIYSerValTYrGIYCyAen 2460  
Db 7132 AACGACATCATMAAGGGAGAACTACAGCTACCGGGCAGTGTGTACCAATGCAT 7191  
Qy 2461 AlaGIYPheArgLeuIleGIYMeSerValArgIleCySGIYngIuAenPHISITrSer 2480  
Db 7192 GCTGGCTTCGGCTCGATGGCATGTCTGTGGCATCTCCAGAGATCATCATCTGTCG 7251  
Qy 2481 GIYTYrThrProPheCyValProIleThrCySGIYHISProGIYAenProValAenGIY 2500  
Db 7252 GGCAAGACCCCTTCTGTCA---ATTACCTGTGACACCCAGGCAACCTGTCAAGGC 7307  
Qy 2501 LeuThrGIYngIYAenGIYPheAenLeuAenAenValValIYVpHeValCYAenProGIY 2520  
Db 7308 CTCATCAGGGTAAACAGATTAACTCAACGATGTGTCAAGTTGTGTGAACCCCTGGG 7367  
Qy 2521 TyrMetAlaGIYngIYAlaAlaArgSerGIYngIYLeuAlaSerGIYngIYTrpSerAspMet 2540  
Db 7368 TATATGGGTAGGGGGCTGTAGTCCAAATGCTGTGGCCAGGGCAATGAGTGCATG 7427  
Qy 2541 LeuProThrCYAArgIleIleAenCYrThrAspProGIYHISngIuAenSerValArg 2560  
Db 7428 CTGCCACCTGCAGATCATCATCACTGTACAGTCTGTGACACCAAGAAATAGTTCGT 7487  
Qy 2561 GlnValHISAlaSerGIYProHISArgPheSerPheGIYThrValSerTYrArgCyS 2580  
Db 7488 CAGGTCCACGGCAGGGGCCACAGTTCAGTTCGGCACACTGTCTTCAACCGGGTC 7547  
Qy 2581 AenHISGIYThrLeuGIYThrProValLeuSerCySGIYngIYAspGIYThrTrp 2600  
Db 7548 AACCAAGGCTTCTACCTCTGTGGCACCCCAAGTGTCAAGTGTGGCCATCCGGGCTCC 7607  
Qy 2601 AspArgProArgProGIYngIYLeuLeu-----ValSerCySGIYHISProGIYSer 2617  
Db 7608 GACCGTCCCGGCCCGCCAGTGTCT--CTGAAGTATGTGTCTGTGGCCATCCGGGCTCC 7666  
Qy 2618 ProProHISerGIYMeSerGIYAspSerTYrThrValAlaValAlaArgTYrSer 2637  
Db 7667 CCGCTCACTCCAGATGTCTGGAGACAGTTACTGTGGAGGAGTGTGTCCGGTACAGC 7726  
Qy 2638 CysIleGIYVpArgThrLeuValGIYAenSerThrArgMetCySGIYLeuAspGIYHIS 2657  
Db 7727 TGCATCGCAAGCGTACTGTGTGGAAACAGACCCGATGTGGGCTGAAATGGAAC 7786  
Qy 2658 TrpThrGIYSerLeuProHISCySerGIYThrSerValGIYValCYSGIYAspProGIY 2677  
Db 7787 TGGATGGCTCCCTCCCTCACTGTGTCAAGAACAGCGTGGAGATTGGCGTGAACCTGG 7846  
Qy 2678 IleProAlaHISGIYIleArgLeuGIYAspSerPheAspProGIYThrValMetArgPhe 2697  
Db 7847 ATCCCGGCTCATGTGATCCGTTGGGGACAGCTTGTATCCAGGACACTGTATGCGCTTC 7906  
Qy 2698 SerCySGIYngIYHISValLeuArgGIYSerSerGIYngIYTrpThrCYGlnAlaAenGIY 2717  
Db 7907 AGCTGTGAAGCTGGCCAGCTGTCCGGGATGTGTAGAGCCGACCTGTCAAGCCAAATGGC 7966  
Qy 2718 SerTrpSerGIYSerGIYngIYValIleSerCySGIYAspProGIYThrPro 2737

Db 7967 TCGTGAAGCGGCTGCAGCCCTAGTGTGAGTATCTTGTGGAAACCTTGGACTCCA 8026  
Qy 2738 SerAenAlaArgValValIlePheSerAspGIYLeuValIlePheSerSerIleValTYrGIY 2757  
Db 8027 AGTATGCTCCGATGTGTGTATGATGGCTGTGTTCCTCAGCTCATGTGTATGAG 8086  
Qy 2758 CybArgIuGIYTYrTYrAlaThrGIYLeuLeuSerArgHISCySerSerValAenGIYThr 2777  
Db 8087 TGGCGGAGAGATCTAGCCACAGGCTGTGTACCCGTCACTCTGGTCAATGTATCC 8146  
Qy 2778 TrpThrGIYSerAspProGIYCyLeuValIleAenCySGIYAspProGIYIleProAla 2797  
Db 8147 TGCACAGGAGTGAACCTGTAGTGTCTGTCTAATACTGTGTGACCTGGGATTCAGGC 8206  
Qy 2798 AenGIYLeuArgLeuGIYAenAspPheArgTYrAenTYrThrValIleThrTYrGIYCyVal 2817  
Db 8207 AATGGCTTCGGCTGGCGAATGACTTCAAGTACAAACATGTGACATACATGTGTCTC 8266  
Qy 2818 ProGIYTYrMetMetGluSerHISArgValSerValLeuSerCyThrTYrAspArgThr 2837  
Db 8267 CTTGGCTATATGATGAGTGCATATAGATATCTGTCTGAGCTGCACCAAGACCGAGCA 8326  
Qy 2838 TrpAenGIYThrIYpProValCYAlaValLeuMetCySGIYpProProLeuIlePro 2857  
Db 8327 TGAATGGAACCAAGCCGCTGTGCAAGCTCTCATGTGCAAGCCACCTCCGCTCATCCC 8386  
Qy 2858 AenGIYVpValValIleGIYSerAspPheMetTrpGIYSerSerValIleThrTYrAlaCyLeu 2877  
Db 8387 AATGGAAAGGTGTGGGTGTGACTTCAATGTGGGGCTCAAGTGTGACTTAAAGCTGTCTG 8446  
Qy 2878 GlnGIYTYrGIYLeuSerLeuProAlaValPheThrCySGIYngIYAsnGIYSerTrpThr 2897  
Db 8447 GAGGGTATACAGCTCTCCCTGCCCGGCTGTTCACCTGTAGGAAATGGTCTGTGACC 8506  
Qy 2898 GIYGIYLeuProGIYngIYpPheProValPheCySGIYAspProGIYValIProSerArgGIY 2917  
Db 8507 GGAAGAGCTGCTCAGTGTTCCTGTGTTCGCGGGATCCTGTGTCCCGTCCGTTGGG 8566  
Qy 2918 ArgArgIuAspArgIYpPheSerTYrArgSerSerValSerPheSerCyHISProPro 2937  
Db 8567 AGGAGAAAGACCCAGGCTTCTCTACAGGTCACTGTCTCTCTCGGCATCCCTCT 8626  
Qy 2938 LeuValLeuValGIYSerProArgArgPheCySGIYngIYAspGIYThrTrpSerGIYThr 2957  
Db 8627 CTGTGTCTGTGGGCTCTCACAGAGTTTGGCAAGTCAATGGAGATAGAGTGCACCC 8686  
Qy 2958 GlnProSerCyAlaAspProThrLeuThrThrCYAlaAspProGIYValIProGIYpHe 2977  
Db 8687 CAGCCAGCTGCATATAGATCCGACCTGACACAGTGTGGGACCTGTGTGTCCACAGTTT 8746  
Qy 2978 GlnIleGIYLeuAenSerGIYngIYTYrGIYngIYSerThrValIlePheArgCyGln 2997  
Db 8747 GGGATACAAACAATTTCCAGGGCTTACAGGTTTGAAGCAACAGTCTTCCGTTCGTCGA 8806  
Qy 2998 IYSGIYTYrLeuLeuGIYngIYSerThrThrArgThrCYLeuProAenLeuThrTrpSer 3017  
Db 8807 AAGGATCACTGTCTTCAGGGCTCCACACAGAGACTGTCTCCAAACCTGTACCTGAAGT 8866  
Qy 3018 GlnTYrProProAspCyValProHISISCyAArgGlnProGIYngIYThrProThrHISAla 3037  
Db 8867 GGAACCCCACTGACTGTGTCCCAACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8926  
Qy 3038 AenValGIYAlaLeuAspLeuProSerMetGIYTYrThrLeuIleThrProAlaArgArg 3057  
Db 8927 AACGTGGGGGCTCGAATTTGCTCTCATAGGCTACACGCTCATATCTCTGCGCAGAGG 8986  
Qy 3058 AlaSerProSerArgValAlaProSerThrAlaProAlaArgArgMetAlaAlaGIYngIY 3077  
Db 8987 GCTTCTCTCCACAGGCTGTCCGAGACAGGCACTGTCAAGGCGGAGCTGTGACAG 9046  
Qy 3078 AlaSerArgProSerAlaIleArgSerGIYProValGIYAspProSerThrLeuProGIY 3097  
Db 9047 GCAAGCGGCCATGTGCTGGAAGTCCGGCCAGTGGAGAGCCATCAACACTCCCGGG 9106

QY 3098 SerHisArgSerProlysePro 3104  
 DB 9107 AGCCACCGCTCACCCAGCCT 9127  
 RESULT 6  
 ID ADH71135 standard; DNA; 10655 BP.  
 XX  
 AC ADH71135;  
 XX  
 DT 25-MAR-2004 (first entry)  
 XX  
 DE Human gene of the invention NOVA4 SEQ ID NO:31.  
 XX  
 KW ds; gene; human; cytosolic; immunomodulator; neuroprotective; nootropic;  
 KW anorectic; antidiabetic; antimicrobial; antipneumatic; gene therapy;  
 KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;  
 KW obesity; diabetes; infectious disease; metabolic syndrome X;  
 KW dyslipidaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003102155-A2.  
 XX  
 PD 11-DEC-2003.  
 XX  
 PF 03-JUN-2003; 2003WO-US017430.  
 XX  
 PR 03-JUN-2002; 2002US-0385120P.  
 PR 04-JUN-2002; 2002US-0385784P.  
 PR 05-JUN-2002; 2002US-0386041P.  
 PR 05-JUN-2002; 2002US-0386047P.  
 PR 06-JUN-2002; 2002US-0386376P.  
 PR 06-JUN-2002; 2002US-0386453P.  
 PR 06-JUN-2002; 2002US-0386864P.  
 PR 06-JUN-2002; 2002US-0387016P.  
 PR 07-JUN-2002; 2002US-0387969P.  
 PR 07-JUN-2002; 2002US-0388161P.  
 PR 07-JUN-2002; 2002US-0388931P.  
 PR 07-JUN-2002; 2002US-0386942P.  
 PR 07-JUN-2002; 2002US-0386971P.  
 PR 07-JUN-2002; 2002US-0387262P.  
 PR 08-JUN-2002; 2002US-0296960P.  
 PR 10-JUN-2002; 2002US-0387400P.  
 PR 10-JUN-2002; 2002US-0387535P.  
 PR 11-JUN-2002; 2002US-0387610P.  
 PR 11-JUN-2002; 2002US-0387625P.  
 PR 11-JUN-2002; 2002US-0387634P.  
 PR 11-JUN-2002; 2002US-0387668P.  
 PR 11-JUN-2002; 2002US-0387696P.  
 PR 11-JUN-2002; 2002US-0387702P.  
 PR 11-JUN-2002; 2002US-0387836P.  
 PR 11-JUN-2002; 2002US-0387859P.  
 PR 12-JUN-2002; 2002US-0387933P.  
 PR 12-JUN-2002; 2002US-0387934P.  
 PR 12-JUN-2002; 2002US-0387960P.  
 PR 12-JUN-2002; 2002US-0388022P.  
 PR 12-JUN-2002; 2002US-0388096P.  
 PR 13-JUN-2002; 2002US-0389123P.  
 PR 14-JUN-2002; 2002US-0389123P.  
 PR 14-JUN-2002; 2002US-0389118P.  
 PR 14-JUN-2002; 2002US-0389120P.  
 PR 14-JUN-2002; 2002US-0389144P.  
 PR 14-JUN-2002; 2002US-0389146P.  
 PR 17-JUN-2002; 2002US-0389729P.  
 PR 17-JUN-2002; 2002US-0389742P.  
 PR 18-JUN-2002; 2002US-0389844P.  
 PR 19-JUN-2002; 2002US-0390069P.  
 PR 19-JUN-2002; 2002US-0390209P.  
 PR 21-JUN-2002; 2002US-0390763P.  
 PR 17-JUL-2002; 2002US-0396706P.  
 PR 06-AUG-2002; 2002US-0401628P.  
 PR 09-AUG-2002; 2002US-0402156P.

PR 09-AUG-2002; 2002US-0402256P.  
 PR 09-AUG-2002; 2002US-0402389P.  
 PR 12-AUG-2002; 2002US-0402786P.  
 PR 12-AUG-2002; 2002US-0402816P.  
 PR 12-AUG-2002; 2002US-0402816P.  
 PR 12-AUG-2002; 2002US-0402821P.  
 PR 12-AUG-2002; 2002US-0402832P.  
 PR 13-AUG-2002; 2002US-0403448P.  
 PR 13-AUG-2002; 2002US-0403459P.  
 PR 13-AUG-2002; 2002US-0403531P.  
 PR 13-AUG-2002; 2002US-0403532P.  
 PR 13-AUG-2002; 2002US-0403563P.  
 PR 13-AUG-2002; 2002US-0403617P.  
 PR 15-AUG-2002; 2002US-0403617P.  
 PR 26-AUG-2002; 2002US-0406182P.  
 PR 26-AUG-2002; 2002US-0406355P.  
 PR 27-AUG-2002; 2002US-0406240P.  
 PR 12-SEP-2002; 2002US-0410084P.  
 PR 20-SEP-2002; 2002US-0412528P.  
 PR 23-SEP-2002; 2002US-0412731P.  
 PR 30-SEP-2002; 2002US-0414801P.  
 PR 30-SEP-2002; 2002US-0414839P.  
 PR 30-SEP-2002; 2002US-0414839P.  
 PR 30-SEP-2002; 2002US-0414954P.  
 PR 09-OCT-2002; 2002US-0417186P.  
 PR 09-OCT-2002; 2002US-0417406P.  
 PR 23-OCT-2002; 2002US-0420639P.  
 PR 28-OCT-2002; 2002US-0421156P.  
 PR 31-OCT-2002; 2002US-0422690P.  
 PR 01-NOV-2002; 2002US-0423130P.  
 PR 05-NOV-2002; 2002US-00423798.  
 PR 05-NOV-2002; 2002US-0423798P.  
 PR 12-NOV-2002; 2002US-0425453P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Alsbrook JP, Alvarez E, Anderson DW, Boldog FL, Casman SJ;  
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K,  
 PI Eutenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X,  
 PI Gusev VY, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR,  
 PI MacLachlan T, Malayanar UM, Mezick AJ, Miller I, Mishra VS;  
 PI Padigan M, Paturajan M, Pena CE, Peyman JA, Raha D, Raetzelli L,  
 PI Rieger DK, Rothenberg ME, Scioe P, Shenoy SG, Shinkets RA,  
 PI Smithson G, Spytek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M,  
 PI Zhong H;  
 XX  
 DR WPI; 2004-081935/08.  
 DR P-PsDB; ADH71136.  
 XX  
 PT New NOVA polypeptides and nucleic acid molecules useful for preventing or  
 PT treating NOVA-associated disorders, e.g. cancer, diabetes, infection or  
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.  
 XX  
 PS Example 4; SEQ ID NO 31; 1680pp; English.  
 XX

CC The invention relates to a novel isolated polypeptide (NOVA). A  
 CC polypeptide of the invention has cytosolic, immunomodulator,  
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and  
 CC antipneumatic activity, and may have a use in gene therapy, and as a  
 CC vaccine. The polypeptides are encoded by NOVA polynucleotides comprising  
 CC any of the 303 fully defined nucleotide sequences given in the  
 CC specification. The polypeptide is useful in the manufacture of a  
 CC medicament for treating a syndrome associated with a human disease. The  
 CC polypeptide, polynucleotide and antibody are useful in diagnosing.  
 CC treating or preventing NOVA-associated disorders, e.g. cancer, cachexia,  
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious  
 CC diseases, metabolic syndrome X or dyslipidaemia. The nucleic acids are  
 CC further used as hybridisation probes, in chromosome mapping, tissue  
 CC typing, preventive medicine, and pharmacogenomics. The present sequence  
 CC encodes a NOVA polypeptide of the invention.  
 XX

SQ Sequence 10655 BP; 2255 A; 3236 C; 2879 G; 2285 T; 0 U; 0 Other;

Alignment Scores:



```
Db 2686 CCCAGTGTGGAAGCTCTGTGTGTGCTTCATTCAGAGCTCCAGGGACCATCTTGTG 2745
Qy 697 ProGlyPheProAspPheTyrProAsnAsnLeuAsnQyStrTrpIleIleGluThrSer 716
Db 2746 CCAGGTTTCCCTGACCTTCTTACCCCAACAATTGAATCAGCTGATTAATGAACAATCT 2805
Qy 717 HisGlyLysGlyValPhePheThrPheHisStrPheHisIleuGlyuSerGlyHisAspTyr 736
Db 2806 CATGGCAAGGGGTGTCTTCTTCACTTCCACACCTTCCACTGGAAAGTGGCCATGACTAC 2865
Qy 737 LeuLeuIleThrGluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArg 756
Db 2866 CTCCCATCACTGAGAACGGCAGCTTCCAGCCCTGAGGAGCAGCTAATGATCTCGG 2925
Qy 757 LeuProAlaProIleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIle 776
Db 2926 CTGCGACGCTCCCATCAAGCGCTGGGCTGTATGGCAACTTCACTGCCAGGTCCGCTTCATC 2985
Qy 777 SerAspPheSerMetSerTyrGlyGlyPheAsnIleThrPheSerGlyTyrAspLeuGlu 796
Db 2986 TCTGATTTCTCCATGTGATGTAAGATTCAAATCACTTCTTCAGATACGACTTGGAG 3045
Qy 797 ProGlyGluGluProGlyValProAlaTyrSerIleArgLysGlyLeuGlnPheGlyVal 816
Db 3046 CCCGTGAGAGAGCCCGAGGTCCAGCTCAGCATCCGGAAGGGGCTTGAGTTGGCGTG 3105
Qy 817 GlyAspThrLeuThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIle 836
Db 3106 GGGCAACCTTGACCTTCTCTGCTTCCCGGATACCGGTGAGGGGACCCGCCCATC 3165
Qy 837 ThrCysLeuGlyGlyArgArgLeuTyrSerSerProLeuProArgCysValAlaGlu 856
Db 3166 ACGGCTGTGGGGGCAACGCGCTGTGGAGCTCGCTCGCAAGTGTGTGTGCTGAG 3225
Qy 857 CysGlyAsnSerValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsn 876
Db 3226 TGTGGGAATTCAGTCACAGGACCTCAGGATCACTTGTGTGCTCCCAACTTCTGTGAGAC 3285
Qy 877 TyrAsnAsnAsnHisGlyCysIleTyrSerIleGlnThrGlnProGlyLysGlyIleGln 896
Db 3286 TACAAATACATCATGATGATCATCTACTCATCAAGCCAGCAGGAAAGGAATTCAG 3345
Qy 897 LeuLysAlaIArgAlaPheGlyLeuLeuSerGlyLysValIleuValTyrAspGlyAsn 916
Db 3346 CTAAAGCCAGGGGCAATTCGAATCTCCGAAGAGATGCTCTCAAGGTTATGATGGCAAC 3405
Qy 917 AsnAsnSerAlaArgLeuLeuGlyValPheSerHisSerGluMetIleValThrLeu 936
Db 3406 AACAACTCCGCCCTTGTCTGGAGATTTTAAAGCATTTCTGAGATGATGGGAGTGA 3465
Qy 937 AsnSerThrSerSerSerLeuTyrLeuAspPheIleThrAspAlaGluAsnThrSerLys 956
Db 3466 AACAGACATCCAGCATCTGTGTGATTTCATCATCATGATGCGAAACACAGCAAG 3525
Qy 957 GlyPheGluLeuHisPheSerSerPheGlyLeuLeuIleLysCysGlyLysAspProGlyThrPro 976
Db 3526 GGGTTGAATCTGACCTTTCAGATTTCAGATTCATCAAAATGTGAGAGACCCAGGAACCCC 3585
Qy 977 LysPheGlyTyrLysValHisAspGlyGlyHisPheAlaGlySerSerValSerPheSer 996
Db 3586 AAGTTTGGCTACAAAGTTTCATGATGAGTCAATTTTTCAGAGGAGCTCCGTGCTTCAAC 3645
Qy 997 CysAspProGlyTyrSerLeuArgLysSerGlyLeuLeuCysLeuSerGlyGluArg 1016
Db 3646 TGTGACCTCGATACAGCTTCGCGGGGTGTGAGAGCTGCTGTGCTGATGAGAGAGCC 3705
Qy 1017 ArgThrTrpAspArgProLeuProThrCysValAlaGluCysGlyGlyThrValArgGly 1036
Db 3706 CGGACTGGGACCGGCTTGTCCACCTGTGTGCGAGGTGTGGAGGAGCAGTGTGAGAGA 3765
Qy 1037 GluValSerGlyGlnValLeuSerProGlyTyrProAlaProTyrGlyHisAsnLeuAsn 1056
|||
Db 3766 GAGGTGTGGGGGAGGAGTGTCTGTACCCGGGATATCCAGCTCCCTATGAAACAATCTCAAC 3825
Qy 1057 CysIleTrpThrIleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPhe 1076
Db 3826 TGCATCTGAGACCATTCAGACAGAGCGGCTGACCATTTGGGCTACACTTCCGTGGTGT 3885
Qy 1077 AspThrGluGluValHisAspValLeuArgIleTrpAspGlyProValGluSerGlyVal 1096
Db 3886 GACACAGAGAGAGGTTCACACAGCTGTGCGCATCTGGGATGGGCTGTGGAGAGCGGGT 3945
Qy 1097 LeuLeuLysGlyLeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSer 1116
Db 3946 CTGCTGAAGAGCTGATGTGAGCCCGGCTTGGCCCAAGACCTGCATATGACCTTCAACTCG 4005
Qy 1117 ValValLeuGlnPheSerThrAspPheThrSerLysGlnGlyPheAlaIleGlnPhe 1136
Db 4006 GTGCTCTTCAGATTGACGACTGACTTCTTACACAGCAAGAGGCTTTCCTCAATTT 4065
Qy 1137 SerValSerThrAlaThrSerCysAsnAspProGlyIleProGlnAsnGlySerArgSer 1156
Db 4066 TCAAGTTCCACAGCAAGCTCTGCAATGACCTGGGATCCGCAAAATGGAGTGGAGT 4125
Qy 1157 GlyAspSerTrpGlyAlaGlyAspSerThrValPheGlnCysAspProGlyTyrTrpAlaLeu 1176
Db 4126 GGTGACAGTTGGGAAGCCGGCGCATCTCCACAGTTCAGGTGTGACCTTGGCTACGCGCTG 4185
Qy 1177 GlnGlySerAlaGluIleSerCysValLysIleGluAsnArgPhePheTrpGlnProSer 1196
Db 4186 CAGGGAATGTGACAGATCAGCTGTGTGAATGAGTGAAGACAGATTCCTTCTGCGAGCCAGC 4245
Qy 1197 ProProTrpCysIleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeu 1216
Db 4246 CGGCCAATCATGATGCTCTCTGGGGGAGAACCTGCAGAGGACCATCTGAGTCAATCTTC 4305
Qy 1217 SerProAsnTyrProGluProTyrProProGlyLysGlyCysAspTrpLysValThrVal 1236
Db 4306 TCACCAATATACCGAAGACCTTACCCGCAAGGAGTGTACTGGAAGTGAACGCTC 4365
Qy 1237 SerProAspTyrValIleAlaLeuVal---PheAsnIlePheAsnLeuGluProGlyTyr 1255
Db 4366 TCACCAAGCTACGTCATCGCCTGTGTATGTGTCCCAAGCTTAACTGGAGCTGCGCAT 4425
Qy 1256 AspPheLeuHisIleTyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyr 1275
Db 4426 GACTTCTCCATATCTAGAGAGAGCGGACTCTTCAAGCCCTTCATATGAGAGGCTTGTAT 4485
Qy 1276 GlySerGlnLeuProGlyArgIleGlySerSerSerAsnSerLeuPheLeuAlaPheArg 1295
Db 4486 GGGTCCAGCTCCAGGCGCCGATGAAAGCAGAGCAAGAGCTTCTCGCTCGCTCGC 4545
Qy 1296 SerAspAlaSerValSerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGlu 1315
Db 4546 AGCCATGATCTGTGACAAAGCTGTGCTGTATGATGACTTCCAGAAAACCCCGGGAG 4605
Qy 1316 SerCysPheAspProGlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeu 1335
Db 4606 TCATGTTTATCTGCTGTTCCATCAAGAAAGGCAACGGGTGGGCTCGACCTGAACTG 4665
Qy 1336 GlySerSerValThrTyrTyrCysHisGlyGlyTyrGluValGluGlyThrSerThrLeu 1355
Db 4666 GGGTCTCTCCGTACACTACTCTGCAACGGGGGCTACGAAAGTTGAGGAGCACTCGACCTG 4725
Qy 1356 SerCysIleLeuGlyProAspGlyLysProValTrpAsnAsnProArgProValCysThr 1375
Db 4726 AGCTGATCTTGGGGCTGTATGGGAAGCCGTGTGAAACAATCCCGGCGAGTGTCAACA 4785
Qy 1376 AlaProCysGlyGlyGlnTyrValGlySerAspGlyValValLeuSerProAsnTyrPro 1395
Db 4786 GCCCCTGTGGGAGACAGTATGTGGTTGAGACGAGTGTGTGCTTGTCCCAACTACCC 4845
Qy 1396 GluAsnTyrThrSerGlyGlnIleCysLeuTyrPheValThrValProLysAspTyr--- 1414
|||
Db 4846 CAGAACTACACAGTGAAGATGCTGTGTATTTTGTATCTGTGCCCAAGGACTATGAT 4905
```

|    |      |   |      |    |      |   |      |
|----|------|---|------|----|------|---|------|
| Qy | 1415 | ValValPheGlyGlnPheAlaPhePheHisThrAlaLeuAsnAspValValGluValHis    | 1434 | Qy | 1775 | GluProAsnHisAspTyrIleGluIleArgAsnGlyProTyrGluThrSerArgMetMet    | 1794 |
| Db | 4906 | GTGGTGTGGCCAGTTCGCTCTTTCACAGCGCCCTCAACAGAGTGTGAGAGTTAC          | 4965 | Db | 5886 | GAGCCCAACAGACTCATAGAAATCCGAAATGCCCCCTATGAGACAGCGCATGATG         | 6045 |
| Qy | 1435 | AspGlyHisSerGlnHisSerArgLeuLeuSerLeuSerGlySerHisThrGlyGlu       | 1454 | Qy | 1795 | GlyArgPheSerGlySerGluLeuProSerSerLeuLeuSerHisIleValThrThr       | 1814 |
| Db | 4966 | GACGGCCACAGCCAGCACTCGCGGTCTCTAGCTCCCTCGGGCTCCCATACAGGTGA        | 5025 | Db | 6046 | GGAAGATTCAAGTGAACGAGCTTCCAAAGCTCCCTCTCCAGCTCCCAAGAACCAAC        | 6105 |
| Qy | 1455 | SerLeuProLeuAlaThrSerAsnGlnValLeuIleValPheSerAlaValGlyLeuAla    | 1474 | Qy | 1815 | ValTyrPheHisSerAspHisSerGlnAsnArgProGlyThelysLeuGluTyrGlnAla    | 1834 |
| Db | 5026 | TCAGTCCCTTGGCCACCTCCAAATCAAGTTCAATTAGTTCAACGCGCAAGGCGCTCGCA     | 5085 | Db | 6106 | GTAATTTTCCACAGCCAGCACTCCCAAGATTCGCGCAGATTCAGCTGAGTATCAGGCC      | 6165 |
| Qy | 1475 | ProAlaArgGlyPheHisPheValTyrGlnAlaValProArgThrSerAlaThrGluCys    | 1494 | Qy | 1835 | TyrGluLeuGlnGlyCysProAspProGlyProPheAlaAsnGlyIleValArgGlyVala   | 1854 |
| Db | 5086 | CCAGCCAGAGGCTTCCACTTGTCTTACCAAGCGGTTCTCGAACAGCGCCAGCGAGTGC      | 5145 | Db | 6166 | TATGAATTTCAAGAGTCCCAAGCCCAAGAGCCCTTTCCATGGAATTGGAGGGAGCT        | 6225 |
| Qy | 1495 | SerSerValProGluProArgTyrGlyValArgLeuGlySerAspPheSerValGlyAla    | 1514 | Qy | 1855 | GlyTyrAsnValGlyGlnSerValThrPheGluCysLeuProGlyTyrGlnLeuThrGly    | 1874 |
| Db | 5146 | AGCTGTGGCCGGAACCCCGTTATGGCAAGGCGGCACTGCTCGGTGGGGGCC             | 5205 | Db | 6226 | GGTCAACAGTGGGACATCATGTGACCTTGAAGTGCCTCCGGGGGTATCAATTGACTGGC     | 6285 |
| Qy | 1515 | IleValArgPheGluCysAsnSerGlyTyrAlaLeuGlnGlySerProGluIleGluCys    | 1534 | Qy | 1875 | HisProValLeuThrCysGlnHisIleGlyThrAsnArgAsnThrAspHisProLeuProlys | 1894 |
| Db | 5206 | ATCGTCCGCTTGAATGCACCTCCGCGCTATGCTCCGCAAGGGGTGCCAGAGATCGAGTGC    | 5265 | Db | 6286 | CACCTGTCTTCAAGTGTCAACATGGCACCAACCGGAACGTGGGACCAACCCCTGCCCCAAG   | 6345 |
| Qy | 1535 | LeuProValProGlyValAlaLeuAlaGlnThrAsnValSerAlaProThrCysValValPro | 1554 | Qy | 1895 | CysGluValProCysGlyGlyAsnIleThrSerSerAsnGlyThrValTyrSerProGly    | 1914 |
| Db | 5266 | CTCCCTGTGCTTGAGGCTTGGCCCAAGAGATGTCTCAGCGCCACAGCTGTGTGTGCGG      | 5325 | Db | 6346 | TGTGAATCCTTGTGGCGGGAAACATCATCTTCTTCAACGCACTGTGTACTCCCGGGG       | 6405 |
| Qy | 1555 | CysGlyGlyAsnLeuThrGluArgArgGlyThrIleLeuSerProGlyPheProGluPro    | 1574 | Qy | 1915 | PheProSerProTyrSerSerSerSerGlnAspCysValTyrPheIleThrValProIleGly | 1934 |
| Db | 5326 | TGTGAGGCAACCTCACAGAGCGCAGGGGCAACATCCTGTCCCTGGCTTCCAGAGCGG       | 5385 | Db | 6406 | TTCCTAGCCCGATCATCTCAAGCTCCAGAGCTGTGTGTGGCTGATCACCGTGCCATTGGC    | 6465 |
| Qy | 1575 | TyrLeuAsnSerLeuAsnCysValTyrPheIleValValProGluGlyAlaGlyTyrIleGln | 1594 | Qy | 1935 | HisGlyValArgLeuAsnLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThr    | 1954 |
| Db | 5386 | TACTTCAACAGCTCAACTGTGTGTGAGATCGTGTGCTCCCGAAGGGCGTGCATCAG        | 5445 | Db | 6466 | CATGGCTCTCCGCTCAACCTCAGCTGTCTCAGACAGAGCCCTCTGGAATTCATCATCC      | 6525 |
| Qy | 1595 | IleGlnValValSerPheValThrGluGlnAsnThrAspSerLeuGluValPheAspGly    | 1614 | Qy | 1955 | IleThrAspGlyProGlnGlnThrAlaProArgLeuGlyValPheThrArgSerMetAla    | 1974 |
| Db | 5446 | ATCCAAAGTTGCAAGTTGTGACAGAGCAAGACTGCGACTGCGAAGATTTGATGTGT        | 5505 | Db | 6526 | ATCTGGATGGCCACAGCAACAGCAACAGCGCTCGGCTTCTTCAACCCGAGCATGGCC       | 6585 |
| Qy | 1615 | AlaAspAsnThrValThrMetLeuGlySerPheSerGlyThrThrValProAlaLeuLeu    | 1634 | Qy | 1975 | LysLeuThrValGlnSerSerSerSerAsnGlnValLeuLeuLysPheHisArgAspAlaAla | 1994 |
| Db | 5506 | GCAATTAACACTGATACCATGCTGGGAGTTTCTCAGGAACAACTGTGCCCTTCTG         | 5565 | Db | 6586 | AAGAAACAGTGCAGAGTTTATCATCAACAGAGCTCTGCTCAAGTTCCACCGATGAGCACC    | 6645 |
| Qy | 1635 | AsnSerThrSerAsnGlnLeuTyrLeuHisPheTyrSerAspIleSerValSerAlaAla    | 1654 | Qy | 1995 | ThrGlyGlyIlePheAlaIleAlaPheSerAlaTyrProLeuThrLysCysProProPro    | 2014 |
| Db | 5566 | AACAGCACTTCCACACAGCTTACCTTCAATTTCTACATCAGATATCAGGGATCTGCAGCT    | 5625 | Db | 6646 | ACAGGGGGGATCTTCCGCATAGCTTCTCCGCTTATCAGCTCAACAAATGCCCTCTCC       | 6705 |
| Qy | 1655 | GlyPheHisLeuGluTyrTyrLeuThrValGlyLeuSerSerCysProGluProAlaValPro | 1674 | Qy | 2015 | ThrIleLeuProAsnAlaGluValValThrGluAsnGluGluPheAsnIleGlyAspIle    | 2034 |
| Db | 5626 | GCGTTCCACTGGAGTACAAACGATGGGCTGTGAGCAGTTGTCCGGAACCTGCTGTCCC      | 5685 | Db | 6706 | ACCATCTCTCCCAACGCGCAAGTGTCAAGAGATGAAGAAATTCATATATGATGATCATC     | 6765 |
| Qy | 1675 | SerAsnGlyValValSerThrGlyGluArgTyrLeuValAsnAspValValSerPheGlnCys | 1694 | Qy | 2035 | ValArgTyrArgCysLeuProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLys    | 2054 |
| Db | 5686 | AGTAACGGGGTGAAGACTGGGAGCGCTACTTGTGATATATATGTGTCTTTTCAAGTGT      | 5745 | Db | 6766 | GTAAGCTACAGATGCTCCCTGCTTACTTGTGGGGAATGAATTCAGACTGGCAA           | 6825 |
| Qy | 1695 | GluProGlyTyrAlaLeuGlnGlyValHisAlaHisIleSerCysMetProGlyThrValArg | 1714 | Qy | 2055 | LeuGlyThrTyrLeuGlnPheGluGlyProProProIleCysGluValHisCysProThr    | 2074 |
| Db | 5746 | GAGCCGGGATATGCTCTCCAGGGGCCAGCGCCACATCTCTGCATGCCCGGAACAGTGGG     | 5805 | Db | 6826 | CTTGGAACTTCACTGCAAGTTTGAAGGACCAACCCCGCATATAGTAAAGTGCATGTCCAA    | 6885 |
| Qy | 1715 | ArgThrPheThrProProProLeuCysIleAlaGlnCysGlyGlyThrValGluGluMet    | 1734 | Qy | 2075 | AsnGluLeuLeuThrAspSerThrGlnValIleLeuSerGlnSerTyrProGlySerTyr    | 2094 |
| Db | 5806 | CGATGGAACATACCTCTCTCCACTGTATGTCACAGTGTGGGGAACAGTGGAGAAATG       | 5865 | Db | 6886 | AATGAGCTTCTGACAGACTCACAGGGGTATCTTGAGCCAGACTACCTCGGAAGCTAT       | 6945 |
| Qy | 1735 | GluGlyValIleLeuSerProGlyPheProGlyAsnTyrProSerAsnMetAspCysSer    | 1754 | Qy | 2095 | ProGlnPheGlnThrCysSerTyrPheValArgValGluProAspTyrAsnIleSerLeu    | 2114 |
| Db | 5866 | GAGGGGGTGAATCCTGAGGCCCGGCTTCCAGGCAACTACCCCAATACATGATGCTCC       | 5925 | Db | 6946 | CCCAAGTTCCAGACGCTGCTTGTGGCTGTAGAGTGAAGCCGACATATACATCTCCTC       | 7005 |
| Qy | 1755 | TyrPheIleAlaLeuProValGlyPheGlyValHisIleGlnPheLeuAsnPheSerThr    | 1774 | Qy | 2115 | ThrValGluTyrPheLeuSerGlyLysGlnTyrAspGluPheGluIlePheAspGlyPro    | 2134 |
| Db | 5926 | TGGAATAATAGCACTGCGCGTGGGCTTGGAGCTCAATCCAGTTCTGAACTTCTCACCC      | 5985 | Db | 7006 | ACAGTGAATGATCTTCTCAGCGAGAAAGCAATATGATGATGATTTTGTATGTGTCA        | 7065 |
| Qy |      |   |      | Qy | 2135 | SerGlyGlnSerProLeuLeuLysAlaLeuSerGlyAsnTyrSerAlaProLeuIleVal    | 2154 |

|   |      |   |      |   |      |  |      |
|---|------|---|------|---|------|--|------|
| D | 7066 | TACGACAGAGTCTCTGCTGAAAGCCCTCAGTGGAAATTACTCAGCTCCCTCGATTTC         | 7125 | D | 8146 | GGCAACCCCTGTCACAGCGCTCACTCAGGGGTAAACCACTTTAACTCAACAGTGTGTCACG      | 8205 |
| Q | 2155 | ThrSerSerSerAsnSerValIleuArgIlePheSerSerAspHisAlaIleuAsnArgIle    | 2174 | Q | 2515 | PheValCysAsnProGlyIleuMetAlaGluGlyAlaAlaArgSerGlnCysLeuAlaSer      | 2534 |
| D | 7126 | ACCAGCTCAAGCAACTCTGTGTACTGCTGCTGTATCATGATCAGCCTTACATTCGAG         | 7185 | D | 8206 | TTTGTGTGCAACCTTGGGTATATAGCTGAGGGGGCTGTAGGTCCTCAATAGCTGGCCAG        | 8265 |
| Q | 2175 | GlyPheIleArgIlePheSerSerAlaProIleCysSerLeuProArgAlaProLeuHisGly   | 2194 | Q | 2535 | GlyGlnIlePheSerAspMetLeuProThrCysArgIleIleAsnCysPheThrAspProGlyHis | 2554 |
| D | 7186 | GGCTTCAGATCCGCTATTACAGCCCTTACTGACACCTGGCCAGGGCTTCACTCATGCGC       | 7245 | D | 8266 | GGGCATAGAGATGACATGCTGCCACCTCCAGATATATCACTGATCAAGATCCTGGACAC        | 8325 |
| Q | 2195 | PheIleLeuGlyGlnIleThrSerThrGlnProGlyGlySerIleHisPheGlyCysAsnAla   | 2214 | Q | 2555 | GlnGluAsnSerValArgGlnValHisAlaSerGlyProHisArgPheSerPheGlyThr       | 2574 |
| D | 7246 | TTCACTCTAGAGCCAGACGACGACCCAGCCGGGGCTCATCTGACTTTGGCTGCAAGCC        | 7305 | D | 8326 | CAAGAAATAAGTTGTGTGACAGGTCCAGCCAGCGGCCGACAGGTTCAGCTTCGCAAC          | 8385 |
| Q | 2215 | GlyIleArgLeuValGlyHisSerMetAlaIleCysThrArgHisProGlnGlyThrHis      | 2234 | Q | 2575 | ThrValSerIleArgCysAsnHisGlyPheIleuLeuGlyIleThrProValLeuSerCys      | 2594 |
| D | 7306 | GGCTACCGCTGTGTGGAGACAGAGTGGCATCTGTACCCGGCACCCTCAGGGCTACAC         | 7365 | D | 8386 | ACTGTGTCTTACCGGTGCAACAGGCTTCTACCTCTGGGGACCCCACTGCTCAGCTGC          | 8445 |
| Q | 2235 | LeuIlePheSerGluAlaIleProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaPro   | 2254 | Q | 2595 | GlnGlyAspGlyThrIlePheAspArgProArgProGlnCysLeuLeuValSerCysGlyHis    | 2614 |
| D | 7366 | CTGTGAGAGGAGAGCATCTCTCTGTACAGCTCTTCTGTGGAGCTTCTGAGGGCCCC          | 7425 | D | 8446 | CAGGAGATGGACATGGGACCGTCCCGCCCCCAGTGTCTGTGTCTGTGTCTGTGCAT           | 8505 |
| Q | 2255 | LysAsnGlyMetValPheGlyValGlyIleuThrValGlyIleuValAlaValIleuSerCys   | 2274 | Q | 2615 | ProGlySerProProHisSerGlnMetSerGlyAspSerIleThrValGlyAlaValAla       | 2634 |
| D | 7426 | AAGATGGAAATGCTGTTGGCAGAGAGTACACAGTGGGACCAAGGCGCATGTACAGCTGC       | 7485 | D | 8506 | CCGGCTCCCGCTCACTCCAGATGTCTGGAGACATTTATCTGTGGAGCACTGTGTG            | 8565 |
| Q | 2275 | SerGluGlyIleuHisLeuGlnAlaGlyAlaGluAlaThrAlaGluCysLeuAspThrGly     | 2294 | Q | 2635 | ArgIleSerCysGlyIleGlyLysArgIleuValGlyAsnSerThrArgMetCysGlyLeu      | 2654 |
| D | 7486 | AGTAAAGCTTACCACTTCACAGGAGGCGTGAAGGCTGAGGCTGAGAGTCTGGACACAGG       | 7545 | D | 8566 | CGTACAGCTGATCGGCAAGGATCTGTGTGGGAAACAGACCCCATGTGTGGCTGC             | 8625 |
| Q | 2295 | LeuIlePheSerAsnArgAsnValProProGlnCysValProValThrCysProAspValSer   | 2314 | Q | 2655 | AspGlyHisIleThrPheGlySerLeuProHisCysSerGlyThrSerValGlyValCysGly    | 2674 |
| D | 7546 | CTATGGAGCAACCCCAATGTCCACACAGTGTCTGTGACCTGTCTGTATGTCAAGT           | 7605 | D | 8626 | GATGGACACTGAGCTGAGCTCTCTCTCTCACTGCTCAGGACACAGTGGAGTTCGGT           | 8685 |
| Q | 2315 | SerIleSerValGluHisGlyArgIleuThrArgLeuIlePheGluThrGlnIleuArgValIle | 2334 | Q | 2675 | AspProGlyIleProAlaHisGlyIleArgLeuGlyAspSerPheAspProGlyIleThrVal    | 2694 |
| D | 7606 | AGCATACGCTGAGGACATGGCCGATGAGGCTTATCTTTGAGACACAGTATCAAGTTCAG       | 7665 | D | 8686 | GACCTGGGATCCCGGCTCAGTGCATCGCATCCGTTTGGGGACACCTTTGATCAGGACGTG       | 8745 |
| Q | 2335 | AlaGlnLeuMetLeuIleCysAspProGlyIleuThrIleuThrGlnIleuArgValIle      | 2354 | Q | 2695 | MetArgPheSerCysGluAlaGlyHisIleuValArgIleuSerSerGluArgIleuThrCysGln | 2714 |
| D | 7666 | GCCAGCTGATGTCTCATCTGTGACCTCGGCTACCTACTACTGCTGCTGCTGCTGCTGCTG      | 7725 | D | 8746 | ATGCGCTTCACTGATGAGGTGCGCACATGCTCGGGGATGTCAAGCGGACCTGTCTCA          | 8805 |
| Q | 2355 | CysGlnAlaAsnGlyIleuThrPheSerLeuGlyAspSerThrProThrCysArgIleIleSer  | 2374 | Q | 2715 | AlaAsnGlySerIlePheSerGlySerGlnProGluCysGlyValIleSerCysGlyAsnPro    | 2734 |
| D | 7726 | TGTAGGCTCAATGCAATGAGAGCTCGGGGACTTACGCCCACTGCGGCAATCATCTCC         | 7785 | D | 8806 | GCCATGCTGTGTGAGGAGGCTCCGACCTGAGGTGAGTGAATCTTGTGGGAACCT             | 8865 |
| Q | 2375 | CysGlyGluLeuProIleProProAsnGlyHisArgIleGlyIleuThrLeuSerValIleGly  | 2394 | Q | 2735 | GlyThrProSerAsnAlaArgValAlaPheSerAspGlyLeuValPheSerSerIle          | 2754 |
| D | 7786 | TGTGGAGAGCTCCCGATTCCTCCCATGAGCCACCGCATGCGAACACTGTCTGTCTACCGG      | 7845 | D | 8866 | GGGACTCCAAGTATGCCCCGAGTGTGTTCAGTATGCTGTGTCTCAGCTCTATC              | 8925 |
| Q | 2395 | AlaIleThrAlaIlePheSerCysAsnSerGlyIleuThrLeuValGlySerArgValArgGlu  | 2414 | Q | 2755 | ValIleuGlyCysArgGluGlyIleuThrAlaIleThrGlyLeuLeuSerArgHisCysSerVal  | 2774 |
| D | 7846 | GCAACAGCCCATCTTCTCTGCAATCCGATACACACTGTGGGCTCCAGGGTGTCTGAG         | 7905 | D | 8926 | GTTATAGTGCCTGGAAAGATCTACTACGCCAGGCTGTCTGACCGCTCACTGCTCGCTC         | 8985 |
| Q | 2415 | CysMetAlaAsnGlyLeuThrPheSerGlySerGlyValArgCysLeuAlaGlyHisCysGly   | 2434 | Q | 2775 | AsnGlyThrIleThrGlySerAspProGluCysLeuValIleAsnCysGlyAspProGly       | 2794 |
| D | 7906 | TGTATGCTCAATGAGCTCTGAGAGTGTGAGAGTCCGCTGCTGTGCTGAGACACTGTGGG       | 7965 | D | 8986 | AATGTACTCGACAGGACAGTCACTGAGTCTGTCTATTAACGTGTGTGACCTCGG             | 9045 |
| Q | 2435 | ThrProGluProIleValAlaAsnGlyHisIleAsnGlyGluAsnIleuSerIleuArgIleSer | 2454 | Q | 2795 | IleProAlaAsnGlyLeuArgLeuGlyAsnAspPheArgIleuValIleuThrValIleuThr    | 2814 |
| D | 7966 | ACTCCCTGAGCCCATGTGTCAACGAGACATCAATGGGAGAACTTACAGCTACCGGGGAGT      | 8025 | D | 9046 | ATTTCAGACCAATGAGCTTCCGCTGGGCAATGACTTCAAGGTACAAACAACTGTGACATAT      | 9105 |
| Q | 2455 | ValIleValIleuThrCysAsnAlaGlyPheArgLeuIleGlyMetSerValArgIleCysGln  | 2474 | Q | 2815 | GlnCysValProGlyIleuMetMetGluSerHisArgValSerValLeuSerCysThrIle      | 2834 |
| D | 8026 | GTGTGTATCAATGCAATGCTGCTTCCCGCTGATCGGATGTGTGTGCGCATCTGCCAG         | 8085 | D | 9106 | CAGTGTCTCCCTGTCTATATGATGAGTCAATAGAGTATCTGTGTGAGCTGACCAAG           | 9165 |
| Q | 2475 | GlnAspHisIleIleThrSerGlyIleuThrProPheCysValProIleThrCysGlyHisPro  | 2494 | Q | 2835 | AspArgIleThrIlePheGlnGlyThrIleuProValCysIleuAlaLeuMetCysIleuProPro | 2854 |
| D | 8086 | CAGGATCATCATGCTGTGGGCAAGACCTTCTGTGTGTGCTGATTAACCTGTGGACACCCA      | 8145 | D | 9166 | GACCGGACATGAGATGAAACCAAGCCGCTGGAAGCTCTCATGTGCAAGCACTCCG            | 9225 |
| Q | 2495 | GlyAsnProValAsnGlyLeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValIle      | 2514 | Q | 2855 | LeuIleProAsnGlyIleuValAlaGlySerAspPheMetIleGlySerSerValIleuThr     | 2874 |
|   |      |   |      | D | 9226 | CTCATCCCAATGGAGGTGTGTGGGTGTGACTTCACTGTGGGGCTCAAGTGTGACTTAT         | 9285 |

```
QY 2875 AlaCsteuGluglYTYrGlnLeuSerleuProAlaValPheThrCysGluGlyAsnGly 2894
|||
Db 9286 GCCGCTCGAGGGGTCACACCTCTCCCTGCCCGGCTGTTCACTGTGAGGGAATGGG 9345
|||
QY 2895 SerTrpThrGlyGluLeuProGlnCysPheProValPheCysGlyAspProGlyValPro 2914
|||
Db 9346 TCCGAGACCGGAGAGCTGCTCAGTGTTCCTGTGTTCTCGCGGGATCCGCTGCCG 9405
|||
QY 2915 SerArgIlyArgArgGluAspArgGlyPheSerIlyArgSerSerValSerPheSerCys 2934
|||
Db 9406 TCCCGTGGAGGAGAGAGGACCGAGCTTCTCTCAACAGCTCATCTCTCTCTCTCCG 9465
|||
QY 2935 HisProProLeuValLeuValGlySerProArgArgPheCysGlnSerAspGlyThrTrp 2954
|||
Db 9466 CATCCCTCTGCTGCTGCTGCTGCTCTCCACGACGCTTTCAGTCAAGATGGAGCATGG 9525
|||
QY 2955 SerGlyThrGlnProSerCysIleAspProThrLeuThrCysValAspProGlyVal 2974
|||
Db 9526 AGTGGACACCGCCAGCTGATGATCCGACCTGACACAGTGTGGGACCTGCTGTG 9585
|||
QY 2975 ProGlnPheGlyIleGlnAsnAsnSerGlnGlyTYrGlnValGlySerThrValLeuPhe 2994
|||
Db 9586 CCAAGTTTGGGATACAGAACAATTCTCAGGGCTACAGGTTGGAGACACAGTCTCTTC 9645
|||
QY 2995 ArgCysGlnIlyArgIlyTYrLeuLeuGlnGlySerThrThrArgThrCysLeuProAsnLeu 3014
|||
Db 9646 CGTTGTCAAAAGGCTACTGCTTCAGGGCTCCACCGACCGACTGCTCTCCAAACCTG 9705
|||
QY 3015 ThrTrpSerGlyThrProProAspCysValProHisIleCysArgGlnProGluThrPro 3034
|||
Db 9706 ACCGAGATGGAGACCCACCTGATGTGTCCCCACACATGCGAGCGACGAGACCCA 9765
|||
QY 3035 ThrHisAlaAsnValGlyAlaLeuAspLeuProSerMetGlyTYrThrLeuIle-ThrPr 3054
|||
Db 9766 ACCGATCCCAACGTCGGGGCTCTGATTTGCCCTCCATGCGCTACAGCTCATCTACTCC 9825
|||
QY 3054 OAlaArgArgAlaSerProSerArgValAlaProSerThrAlaProAlaArgArgMetAl 3074
|||
Db 9826 TGCAGAGAGGGCTTCTCTCCACAGGAGTGGCTCCAGACCGCACCTCCAAAGCGGATGGC 9885
|||
QY 3074 aaIaGlyAlaIaSerArgProSerAlaTrpArgSerGlyProValGlyAspProSerTh 3094
|||
Db 9886 AGCTGAGCAGGACGAGCGCCCATCTGCTGAGAGTCCGGCCCAATGGAGATCCATCAAC 9945
|||
QY 3094 rLeuProGlySerHisArgSerProIlyPro 3104
|||
Db 9946 ACTGCCCGGAGGACCGCTACCCAGAGCT 9976
|||

RESULT 7
ADH71143
ID ADH71143 standard; DNA; 10466 BP.
AC
AC ADH71143;
DT
DT 25-MAR-2004 (first entry)
XX
DE Human gene of the invention NOVAe SEQ ID NO:39.
XX
XX de; gene; human; cytosolic; immunomodulator; neuroprotective; nootropic;
XX anorectic; antidiabetic; antimicrobial; antidiabetic; gene therapy;
XX cancer; cachexia; Alzheimer's disease; Parkinson's disease;
XX obesity; diabetes; infectious disease; metabolic syndrome X;
XX dyslipidemia.
XX
XX Homo sapiens.
XX
XX OS
XX PN WC0003102155-A2.
XX
XX PD 11-DEC-2003.
XX
XX PF 03-JUN-2003; 2003WC-US017430.
XX
```

```
PR 03-JUN-2002; 2002US-0385120P.
PR 04-JUN-2002; 2002US-0385784P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386047P.
PR 07-JUN-2002; 2002US-0386376P.
PR 08-JUN-2002; 2002US-0386453P.
PR 09-JUN-2002; 2002US-0386453P.
PR 10-JUN-2002; 2002US-0386453P.
PR 11-JUN-2002; 2002US-0386453P.
PR 12-JUN-2002; 2002US-0386453P.
PR 13-JUN-2002; 2002US-0386453P.
PR 14-JUN-2002; 2002US-0386453P.
PR 15-JUN-2002; 2002US-0386453P.
PR 16-JUN-2002; 2002US-0386453P.
PR 17-JUN-2002; 2002US-0386453P.
PR 18-JUN-2002; 2002US-0386453P.
PR 19-JUN-2002; 2002US-0386453P.
PR 20-JUN-2002; 2002US-0386453P.
PR 21-JUN-2002; 2002US-0386453P.
PR 22-JUN-2002; 2002US-0386453P.
PR 23-JUN-2002; 2002US-0386453P.
PR 24-JUN-2002; 2002US-0386453P.
PR 25-JUN-2002; 2002US-0386453P.
PR 26-JUN-2002; 2002US-0386453P.
PR 27-JUN-2002; 2002US-0386453P.
PR 28-JUN-2002; 2002US-0386453P.
PR 29-JUN-2002; 2002US-0386453P.
PR 30-JUN-2002; 2002US-0386453P.
PR 31-JUN-2002; 2002US-0386453P.
PR 01-JUN-2003; 2003US-0423130P.
```

PR 05-NOV-2002; 2002US-00423798.  
 PR 05-NOV-2002; 2002US-0423798P.  
 PR 12-NOV-2002; 2002US-0425453P.  
 XX

(CURA-) CURAGEN CORP.

XX Alabrook JP, Alvarez E, Anderson DW, Boldog FL, Casman SJ;  
 PI Caterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;  
 PI Eitenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;  
 PI Gusev VY, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, Macdougall JR;  
 PI MacLachlan T, Malyankar UM, Mezick AJ, Millet I, Mishra VS;  
 PI Padigan M, Paturajan M, Pena CE, Peyman JA, Raha D, Rasetelli L;  
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkets RA;  
 PI Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;  
 PI Zhong H;  
 XX

DR WPI: 2004-081935/08.  
 DR P-PSDB; ADH71144.

PT New NOXV polypeptides and nucleic acid molecules useful for preventing or  
 PT treating NOXV-associated disorders, e.g. cancer, diabetes, infection or  
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.  
 XX

PS Example 4; SEQ ID NO 39; 1880bp; English.

XX The invention relates to a novel isolated polypeptide (NOXV). A  
 CC polypeptide of the invention has cytosolic, immunomodulator,  
 CC neuroprotective, neurotropic, anorectic, antidiabetic, antimicrobial, and  
 CC antipneumatic activity, and may have a use in gene therapy, and as a  
 CC vaccine. The polypeptides are encoded by NOXV polynucleotides comprising  
 CC any of the 303 fully defined nucleotide sequences given in the  
 CC specification. The polypeptide is useful in the manufacture of a  
 CC medicament for treating a syndrome associated with a human disease. The  
 CC polypeptide, polynucleotide and antibody are useful in diagnosing.  
 CC CC treating or preventing NOXV-associated disorders, e.g. cancer, cachexia,  
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious  
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are  
 CC further used as hybridisation probes, in chromosome mapping, tissue  
 CC typing, preventive medicine, and pharmacogenomics. The present sequence  
 CC encodes a NOXV polypeptide of the invention.  
 XX

SQ Sequence 10466 BP; 2222 A; 3160 C; 2832 G; 2252 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 10466  
 Score: 15917.50 Matches: 2946  
 Percent Similarity: 95.37% Conservative: 21  
 Best Local Similarity: 94.70% Mismatches: 45  
 Query Match: 93.72% Indels: 100  
 DB: 12 Gaps: 11

US-10-016-248-2 (1-3104) X ADH71143 (1-10466)

QY 2 AlaGlyAlaProProProAlaLeuLeuLeuProCysSerLeuIle---SerAspCys--- 19  
 DB 731 TCTGGTTACCGGAGCCAGCCCTCCAGCCCTTATTCAGCAGCAAGAACTGGTGCAC 790  
 QY 20 CysAlaSerAsnGlnIleArgHisSerValGlyValGly-----ProSerGlnLeuVal 36  
 DB 791 TGCACTTCACATCGATGGACCAACCGGACGCGCGGATTGATCCCAATGCCAA-CTC 849  
 QY 37 LysLysGlnIleGlnLeuLysSerArgGlyValLysLeuMetProSerLysAspAsnSer 56  
 DB 850 AAGAGCAATGATGATGAGTCTGCGAGGTGTGAGAGCTGATGCCAGCAAGAAAGCAAGC 909  
 QY 57 GlnLysThrSerValLeuThrGlnValGlyValSerGlnGlyLysAsnMetCysProAsp 76  
 DB 910 CAGAAGACCTCTGTGTACTCAGGTTGGTGTGTCCTCAAGAGCAATATATGTCTCAGAC 969  
 QY 77 ProGlnIleProGlnIleArgGlyLysArgLeuGlySerAspPheArgLeuGlySerVal 96  
 DB 970 CTTGGCATATCCGAAAGGGGCAAAAGACTAGGCTCGATTTCAGGTTAGATCCAGCGTC 1029

QY 97 GlnPheThrCysAsnGlnGlyLysArgPheLeuGlnGlySerLysArgLysThrCysMetLys 116  
 DB 1030 CAGTTCACTGCAACGAGGGCTATGACTGCAAGAGGTCCAGCGGATCACTCTATATGAA 1089  
 QY 117 ValSerAspMetPheAlaAlaTrpSerAspHisArgProValCysArgAlaArgMetCys 136  
 DB 1090 GTTAGGCAATGTTTTCGCGCTCGAGCGCACACAGGCCAGTCTGCCGAGCCGCATGTCT 1149  
 QY 137 AspAlaHisLeuArgGlyProSerGlyLysIleThrSerProAsnPheProIleGlnTyr 156  
 DB 1150 GATGCCACCTTCGAGGCCCTCGGCGATCATCACTCCCAATTTCCCAATTCAGTAT 1209  
 QY 157 AspAsnAsnAlaHisCysValTrpIleIleThrAlaLeuAsnProSerLysValIleLys 176  
 DB 1210 GACAAACAATGACACACTGTGTGATCATCAGCACTCAACCTCCAGGATATTCAG 1269  
 QY 177 LeuAlaPheGlnGluPheAspLeuGlnArgGlyLysArgPheThrValGlyAspGly 196  
 DB 1270 CTGCGCTTTGAGAGAGTTGATTTGAGAGGGGCTATGACACCTGACGGTGGTATGCT 1329  
 QY 197 GlyGlnAspGlyAspGlnLysThrValLeuTyrMetSerGlnAsnAlaCysSerAspSer 216  
 DB 1330 GGTCAAGATGGGACCAAGACAGTCTCTATCATG----- 1365  
 QY 217 ProHisThrProGlySerArgLysProGlnSerMetSerGlyAspIleTrpArgGlnLys 236  
 DB 1366 -----CTGACAGGATCATCGGTCCCGATCATTCACACACCAAT---CATCAATG 1416  
 QY 237 TrpThrValLeuGlnIleCysArgAspLysSerSerAspAlaArgSerGlySerVal 256  
 DB 1417 TGGCTCCCTTCGAG-----ACGATGCGC---AGTGGCAGT--- 1449  
 QY 257 ArgLysSerProLysThrSerAsnAlaValGlnLeuValAlaProGlyThrGlnIleGlu 276  
 DB 1450 -----TCCCTGGATTCAGAGCTTCTTGAAGAGATGAG 1485  
 QY 277 GlnGlySerCysGlyAspProGlyLysProAlaLysThrValArgGlnGlySerArgPhe 296  
 DB 1486 CAGGCACTTCGCGTACCTGATACCTGCAATACCTGATATGACCGGAGGAGAGGCTCCGGTT 1545  
 QY 297 HisHisGlyAspThrIleLysAspPheGluCysGlnProAlaPheGlnLeuValGlyGlnLys 316  
 DB 1546 CGCCACGGTGACACATCAAGTTGATGAGTGCAGCGCCCTTATGACTGCTGGACAGAG 1605  
 QY 317 AlaIleThrCysGlnLysAsnAsnGlnTrpSerAlaLysLysProGlyCysValPheSer 336  
 DB 1606 GCAATCAATGCAACCAAAAGATTAACCAATGTCGCTAAGAAAGCAGGCTCGTGTGCC 1665  
 QY 337 CysPhePheAsnPheThrSerProSerGlyValValLeuSerProAsnTyrProGluAsp 356  
 DB 1666 TGCTTCTTCAACTTACACACCGCTCTGGGGTGTCTCTGCTCCCACTACCCAGAGGAC 1725  
 QY 357 TyrGlnValMetHisLeuHisCysValTrpLeuIleLeuAlaArgProGlnSerArgLysHis 376  
 DB 1726 TATGGCAACACCTTCACTGTGTCTGCTCATCTGCGCAGGCTGAGGCGGCATATCC 1785  
 QY 377 LeuAlaPheAsnAspIleAspValGlnProGlnPheAspPheLeuValIleLysArgGly 396  
 DB 1786 CTGGCTTCAACACATTTGAGTGTGAGGCTCATGATTTGATTTCTGTGTCAATCAAGATGG 1845  
 QY 397 AlaThrAlaGlnAlaProValLeuGlyLysThrPheSerGlyAsnGlnLeuProSerSerIle 416  
 DB 1846 GCCACCGCGGAGGCGCCGCTCGGCACTTCTCAAGAAACAGTTCCTCTCTCCATC 1905  
 QY 417 ThrSerSerGlyHisValAlaArgLeuGlnPheGlnThrAspHisSerThrGlyLysArg 436  
 DB 1906 ACAAGACAGTGGCCACGTCGCTGCTGAGTTCCAGACTGACACATCCACAGGAAAGAG 1965  
 QY 437 GlyPheAsnIleThrPheThrThrPheArgHisAsnGlnCysProAspProGlyValPro 456  
 DB 1966 GCGTTCAATCACTTTTCCAGTGTGATCTCCCAAGATGCGCGGATCTCGGCTTCCA 2025  
 QY 457 ValAsnGlyLysArgPheGlyAspSerLeuGlnLeuGlySerSerIleSerPheLeuCys 476

```
Db 2026 GTAAATGGCAAAACGGTTTGGGGACAGCCTCCAGCTGGGCACTCATCTCTCTGT 2085
Qy 477 AepgluglyPheLeuGlyThrGlnGlySerGluThrIleThrCysValLeuGlyGlu 496
Db 2086 GATAGAGGCTTCCTTGGGACTCAGAGGCTCAGAGACCATCCTCGCTCTGAAAGAGGC 2145
Qy 497 SerValVal1TrpAnserAlaValLeuArgCysGluAlaProCysGlyGlyValLeuThr 516
Db 2146 AGCGTGTCTGAAACAGCGCTGTGTCTGGGTGAAGCTCCCTGGGTGGTCTGACTGACT 2205
Qy 517 SerProSerGlyThrIleLeuSerProGlyTrpProGlyPheTyrGlyAspAlaLeuSer 536
Db 2206 TCGCCAGCGGACCATCTCTCTCCGGCTGGCTGGCTCTTCAAGAGAGCTTGAAC 2265
Qy 537 CysAlaTrpVal1IleGluAlaGlnProGlyTyrProIleValIleThrPheAspArgPhe 556
Db 2266 TGTGCTGTGGTGAATGAGGCCAGCGCATACCCCATCAAAATCACCTTGACAGGTTTC 2325
Qy 557 LysThrGluValAsnTyrAspThrLeuGluValArgAspGlyArgThrTyrSerAlaPro 576
Db 2326 AAAACCGAGGTCACTATGACACCTGGAGATAGCGATGGCGGACTTACTCAGCGCC 2385
Qy 577 LeuIleGlyVal1TyrIleGlyThrGlnVal1ProGlnPheLeuIleSerThrSerAsnTyr 596
Db 2386 TTGATCCGGGCTTACACAGGAGCCAGGTTCCCATGTTCTCATCAGACCCAGCATAC 2445
Qy 597 LeuTyrLeuLeuPheSerThrAspLysSerHisSerAspIleGlyPheGlnLeuArgTyr 616
Db 2446 CTCTACCTCTCTCTCTACCGACAGAGTCACTCGGACATCGGCTTCCAGCTCCGCTAT 2505
Qy 617 GluThrIleThrLeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGln 636
Db 2506 GAGGCTTAAACATGACGTGACACCATCTGTGAATCCAGGAATCCAGTAAATGAGCAG 2565
Qy 637 ArgIleGlyAsnAspPheTyrVal1GlyAlaLeuVal1ThrPheSerCysAspSerGlyTyr 656
Db 2566 CGTATGAGGAATGACTTCTACGTGGGCGCGCTGGTGACTTCACTGTGACTCGGGCTAC 2625
Qy 657 ThrLeuSerAspGlyGluProLeuGlnCysGluProAsnPheGlnTrpSerArgAlaLeu 676
Db 2626 ACATTAAAGTAGCGGGGAGCCTCTGAGGTGAGCCCAACTTCCAGTGGAGCGGGCCCTG 2685
Qy 677 ProSerCysGluAlaLeuCysGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSer 696
Db 2686 CCCAGTTGTGAAGCTCTGTGGTGGCTTCAATCAAGGCTCCAGTGGAGCATCTTTCG 2745
Qy 697 ProGlyPheProAspPheTyrProAsnAsnLeuAsnCysThrTrpIleIleGluThrSer 716
Db 2746 CCAAGTTTCCCTGACTTCTACCCCAACACTTGAACCTGCACTGGAATTCGAAACATCT 2805
Qy 717 HisGlyGlyGlyVal1PhePheThrPheHisThrPheHisLeuGlnSerGlyVal1AspTyr 736
Db 2806 CATGGCAAGGGGTGTCTTCTTCACTTTCACACTTTCACACTGGAAGGGCCATGACATAC 2865
Qy 737 LeuLeuIleThrGluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArg 756
Db 2866 CTCTCTCATCATGAGAACCGGACGCTTCAACCGAGCCCTGAGAGCGCTTAACTGGAATCTCGG 2925
Qy 757 LeuProAlaProIleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnVal1ArgPheIle 776
Db 2926 CTGCGACGCTCCCATCAGCGCTGGGCTCTATGGCAACTTCACTGCCCGAGTCCGTTATC 2985
Qy 777 SerAspPheSerMetSerTyrGluGlyPheAsnIleThrPheSerGlyTyrAspLeuGlu 796
Db 2986 TCTGATTTCTTCATGTCTATGAGATCAATCAATCACTTCTCAGAGTACGACTTGAG 3045
Qy 797 ProCysGluGluProGluVal1ProAlaTyrSerIleArgGlyGlyLeuGlnPheGlyVal 816
Db 3046 CCCGTGAGAGGCCCGAGGTCCTCAGCTTACAGATCCGGAAAGGGCTTGCAGTTTGGGCTG 3105
Qy 817 GlyAspThrLeuThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIle 836
Db 836 GAGCAGACATTGACCTTCTCTGCTTCCCGGGATACCGTGTGAGAGGCGCCGATC 3165
Qy 837 ThrCysLeuGlyGlyAlaArgArgLeuTrpSerSerProLeuProAlaGlyCysVal1AlaGlu 856
Db 3166 ACGTGCCTGGGGGCAACCGCGCTGTGAGCTCGCTTGCCTGCGCAAGGTGTGTGCGAG 3225
Qy 857 CysGlyAsnSerVal1ThrGlyThrGlnGly1ThrLeuLeuSerProAsnPheProValAsn 876
Db 3226 TGTGGGAATTCAGTCAAGGACCTCAGAGTACTTGTGTGTCGCCCAACTTCTCTGTAC 3285
Qy 877 TyrAsnAsnAsnHisGlyCysAlaIleTyrSerIleGlnThrGlnProGlyTyrGlyIleGln 896
Db 3286 TACAAATAACAATCATGATCATCTACTCATCCATCCAGACCCAGCAGGAAAGGAAATTGAG 3345
Qy 897 LeuLeuAlaArgAlaPheGluLeuSerGlnGlyAspVal1LeuVal1TyrAspGlyAsn 916
Db 3346 CTGAAGCCAGGGGATTCGAATCTCTCGAAGAGAGTGTCTTCAAGGTTTATGATGGCAGAC 3405
Qy 917 AsnAsnSerAlaArgLeuLeuGluVal1PheSerHisSerGluMetMetGlyVal1ThrLeu 936
Db 3406 AACACTCCGCGCTGTGTGTGGAGTTTATGACCATTCATGATGATGGGGGTGACTTGTG 3465
Qy 937 AsnSerThrSerSerSerLeuTrpLeuAspPheIleThrAspAlaGluAsnThrSerLys 956
Db 3466 AACAGCATCCAGACATCTGTGGCTTGAATTCATCATCTGATGCTGAATAACACAGCAG 3525
Qy 957 GlyPheGluLeuHisPheSerSerPheGluLeuIleLysCysGluAspProGlyTyrPro 976
Db 3526 GCGTTTAACTGACATTTTCCAGTTTGAATCATCAATCAATGATGAGACCCAGAACCCCC 3585
Qy 977 LysPheGlyTyrLysVal1HisAspGluGlyHisPheAlaGlySerSerVal1SerPheSer 996
Db 3586 AAGTTTGGCTTAAAGGTTCAAGATGAAGAGTCAATTTTGACAGAGAGCTCCGTGTCTTACG 3645
Qy 997 CysAspProGlyTyrSerLeuAspArgLysSerGluGluLeuLeuLeuLeuLeuLeuLeu 1016
Db 3646 TGTGACCTCGATACAGCCCTCGGGGTATGAGAGAGCTGTGTGTGATGAGAGAGCGC 3705
Qy 1017 ArgThrTrpAspArgProLeuProThrCysVal1AlaGluCysGlyGlyThrVal1ArgGly 1036
Db 3706 CGGACCTGGAGACCGGCTCTGCCACCTGTGTGCCAGGTGAGAGGAGCAGAGAGAGA 3765
Qy 1037 GluValSerGlyGlnVal1LeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsn 1056
Db 3766 GAGGTGTGGGGAGGAGGTGTGTACCCGGGTATTCAGCTCCATGAAACAATCTTCAAC 3825
Qy 1057 CysIleTrpThrIleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuVal1Phe 1076
Db 3826 TGCATCTGSAACATCCAGAGAGAGCGCGCTGCACCATTTGGCTACACTTCTCGTGT 3885
Qy 1077 AspThrGluGluVal1HisAspVal1LeuArgIleTrpAspGlyProVal1GluSerGlyVal1 1096
Db 3886 GACACAGAGAGGTTTCCAGACGTGTGCGCATCTGGGATGGGCTGTGGAGAGACGGGGTT 3945
Qy 1097 LeuLeuLeuGluLeuLeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSer 1116
Db 3946 CTGCTGAAGAGCTGAGTGGCCGGCCCTGAGCCCAAGGACCTGTATGACACTTCACTCG 4005
Qy 1117 ValValLeuGlnPheSerThrAspPhePheThrSerTyrGlnGlyPheAlaIleGlnPhe 1136
Db 4006 GTGCTCTGAGATTCAGACTGACTTCTTCAACGACAGAGGCTTGTGCAATTCATTT 4065
Qy 1137 SerValSerThrAlaThrSerCysAsnAspProGly1IleProGlnAsnGlySerArgSer 1156
Db 4066 TCAGGTTCCACAGCAACGTCTCTCAATGACCTCGGATCCCGAGATGGAGTCCGAGAT 4125
Qy 1157 GlyAspSerTrpGluAlaGlyAspSerThrVal1PheGlnCysAspProGlyTyrAlaLeu 1176
Db 4126 GGTGACAGTTGGGAAGCGCGGAGCTCACAGTGTCTCAGTGTACCTGTGCTAGCGGCTG 4185
Qy 1177 GlnGlySerAlaGluIleSerCysVal1LysIleGluAsnArgPhePheTrpGlnProSer 1196
Db 4186 CAGGGAAGTCCAGAGATCAGCTGTGTGAAGATGGAAGAACAGGTTCTTCTGCGAGCCAGC 4245
```

|    |      |  |      |
|----|------|--|------|
| QY | 1555 | CYGGIYGIYAAsnLeuThrGluArgArgGlyThrIleLeuSerProGlyPheProGluPro    | 1574 |
| Db | 5326 | TGTGGAGCAACTCTCACAGAGCGCAGGGGACACACTCTGTCCCTGGCTTCCAGAACCG       | 5385 |
| QY | 1575 | TYLeuAsnSerLeuAsnCYsValTrrpySileValValProGluGlyAlaGlyIleGln      | 1594 |
| Db | 5386 | TACCTCAACGCGCTCAACTGTGTGTGGAAAGATCGTGGTCCCGAAGCGCGGATCCAG        | 5445 |
| QY | 1595 | IIAGlnValValSerPheValThrGluGlnAsnTrpAspSerLeuGluValPheAspGly     | 1614 |
| Db | 5446 | ATCAAGATTGTCAAGTTTGTGTGACAGAGAACCTGGAGCTCGCTGGAAAGTATTGAAGGT     | 5505 |
| QY | 1615 | AlaAspAsnThrValThrMetLeuGlySerPheSerGlyThrThrValProAlaLeuLeu     | 1634 |
| Db | 5506 | GCAGATTAACACTGTAAACATGTGTGGAGGATTCTCAGGAACACCGCTGCGCTTCTG        | 5565 |
| QY | 1635 | AsnSerThrSerAsnGlnLeuThrIleuHisPheTyrSerAspIleSerValSerAlaIa     | 1654 |
| Db | 5566 | AACAGACCTCCACMACCACTGACTTCATTCTACAGATATCACGATCTGCAGCT            | 5625 |
| QY | 1655 | GIYPheHisLeuGlyTyrLeuThrValAlGlyLeuSerSerCysProGluProAlaValPro   | 1674 |
| Db | 5626 | GGCTTCCACTTGGAGTACAAACCGTGGAGCTTGAGCAAGTTGTCCGAAACTGTGTGGCC      | 5685 |
| QY | 1675 | SerAsnGlyValAlaLeuThrGlyGluArgTyrLeuValAsnAspValValSerPheGlnCys  | 1694 |
| Db | 5686 | AGTAACGGGGTGAAGACTGCGGACCGCTATCTTGATGATGTGTGGTCTTCCAGTGT         | 5745 |
| QY | 1695 | GIUPProGlyTrrAlaLeuGlnGlyHisAlaHisIleSerCysMetProGlyThrValArg    | 1714 |
| Db | 5746 | GAGCGGGAGATGCGCTCTCCAGGGGCGACGCCACACTCTCCGATGCGCCGAACAGCGCG      | 5805 |
| QY | 1715 | ArgTrrAsnTyrProProProLeuCysIleAlaGlnCysGlyValThrValGluGluMet     | 1734 |
| Db | 5806 | CGATGAACTAACCTCCCTCCACTCTGATTTGCACAGTGTGGGGAACAGTGAAGAGATG       | 5865 |
| QY | 1735 | GluGlyValIleLeuSerProGlyPheProGlyAsnTyrProSerAsnMetAspCysSer     | 1754 |
| Db | 5866 | GAGGGGGAGATCTCGAGCGCGGCTTCCAGGGCACTACCCCACTGAACCTGTCTCC          | 5925 |
| QY | 1755 | TrpIlySileAlaLeuProValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThr    | 1774 |
| Db | 5926 | TGGAAAATGACACTGCCCGGTGGCTTTGGAGCTCACTCAAGTTCTGTGAACCTTCTCAC      | 5985 |
| QY | 1775 | GluProAsnHisAspTrrIleGluIleArgAsnGlyProTyrGluThrSerArgMetMet     | 1794 |
| Db | 5986 | GAGCCCAACCCGACTCATAGAAATCCGGAAATGGCCCTCATGAGCACGCCGATGATG        | 6045 |
| QY | 1795 | GlyArgPheSerGlySerGluLeuProSerSerLeuLeuSerThrSerHisGluThrThr     | 1814 |
| Db | 6046 | GGAAAGATTCAGTGGAAACGAGCTTCACAAAGTCTCTCTCTCCACGCTCCACAGACCAAC     | 6105 |
| QY | 1815 | ValTyrPheHisSerAspHisSerGlnAsnArgProGlyPheTyrLeuGlyTyrGlnIa      | 1834 |
| Db | 6106 | GTTGATTTCACAGACGACCACTCCAGAAATCGGACAGATTCAACTGTGAGATTAAGGCC      | 6165 |
| QY | 1835 | TyrGluLeuGlnGlyCysProAspProGluProPheAlaAsnGlyIleValArgGlyAla     | 1854 |
| Db | 6166 | TATAAACTTCAAGAGTGGCCAGACCCAGAGCCCTTTGGCAATGACATGTGAGGGAGACT      | 6225 |
| QY | 1855 | GlyTyrAsnValGlyGlnSerValThrPheGlyCysGluAspProGlyTyrGlnLeuThrGly  | 1874 |
| Db | 6226 | GGCTAACAACTGGGACATTCAGTGACTTCCAGTGTCTCCGGGGTATCAATTGACTGGC       | 6285 |
| QY | 1875 | HisProValLeuThrCysGlnHisGlyThrAsnArgAsnTrpAspHisAspProLeuProlys  | 1894 |
| Db | 6286 | CACCTGTCTCTCAAGTGCACATGCGACCAACCGGAATCGGACCAACCCCTGCCAAG         | 6345 |
| QY | 1895 | CysGluValAProCysGlyGlyValAsnIleThrSerSerAsnGlyThrValTyrSerProGly | 1914 |
| Db | 6346 | TGTGAAGTCCCTTGTGGCGGAAACATCACTTCTTCCAAAGCGCACTGTGTACTCCCGGGG     | 6405 |
| QY | 1915 | PheProSerProTyrSerSerSerGlnAspCysValTrrPheIleThrValProIleGly     | 1934 |

```
Db 6406 TTCCCTACCCCGTACTCCAGCTCCCGAGCATGTGTCTGCTGATCAGCTGCCATTTGGC 6465
Qy 1935 H18G1Val1ArgLeuAanLeuSerLeuLeuGlnThrGluProSerGlyAspPhe1LeuThr 1954
Db 6466 CATGGCGTCCCGCTCAACCTCAGCTGTGAGACAGAGCCCTCTGGAGATTTCTATCCACC 6525
Qy 1955 I1etPAspGlyProGlnGlnThrAlaProArgLeuGlyVal1PheThrArgSerMetAla 1974
Db 6526 ATCTGGATGGGCGCCACAGCAACAGCCACGCTCGCGCTTTTCAACCCGAGCATGGGCC 6585
Qy 1975 Lyb1y8ThrVal1GlnSerSerSerAang1Val1LeuLeuLysPheHisArgAspAla1a 1994
Db 6586 AAGAAAACAGTGCAGAGTTCACTCCACAGGCTCGCTCAAGTTCCACCGGTGATGCAACC 6645
Qy 1995 ThrGlyGly1LeuPheAla1LeuAlaPheSerAla1TyProLeuThrLysCyProProPro 2014
Db 6646 ACAGGGGGGATCTTCGSCATAGCTTTCTCCGCTTATCCACTCACCMAATGCCCTCTCC 6705
Qy 2015 Thr1LeuProAanAlaGlyVal1Val1ThrGluAang1LupPheAan1LeGlyAsp1Le 2034
Db 6706 ACCATCTCTCCCAACGCGGAAGTGTCAAGAGATGAAGATTCATATAGTGAACATC 6765
Qy 2035 Val1ArgTyArgCyLeuProGly1PheThrLeuVal1GlyAang1Lup1LeuThrCysLys 2054
Db 6766 GTAGGCTACAGATGCTCCCTGGGCTTTTACTTAACTAGGGGAATGAATTTCTACCTGCAA 6825
Qy 2055 LeuGlyThrTyLeuGlnPheGluGlyProProProl1eCyGluVal1HisCyProThr 2074
Db 6826 CTGGAACTACCTGACGTGTGAAGAGACACCCCGATATGTGAAGTCACTGTCACCA 6885
Qy 2075 Aang1LeuLeuThrAspSerThrGlyVal11LeuSerGlnSerTyProGlySerTyPro 2094
Db 6886 AATGAGCTTCTGACAGACTCCACAGGCGTGAATCTTGAGCCAGAGTACCTCGAAGACT 6945
Qy 2095 ProGlnPheGlnThrCysSerTyPheVal1ArgVal1GluProAspTyAsn1LeuSerLeu 2114
Db 6946 CCCAGTTCACAGACTGCTCTTGCTGTGAGAGTGCAGCCGACTAATACATCTCCCTC 7005
Qy 2115 ThrVal1GlyTyPheLeuSerGlyLupGlnTyArgGluPheGlu1LeuPheAspGlyPro 2134
Db 7006 ACAGTGGAGTACTCTCTGAGAGACAAATATGATGTGAGTTTGAATTTGATGTGCTCA 7065
Qy 2135 SerGlyGlnSerProLeuLeuVal1AaLeuSerGlyAanTyArgSerAlaProLeu1LeVal 2154
Db 7066 TCAGGACAGAGTCCCTGCTGAAAGCCCTCACTGGAATTACTCACTGCTTATGCTC 7125
Qy 2155 ThrSerSerSerAanSerVal1TyTyLeuArgTyPheSerSerAspHisAlaTyArgAlaGly 2174
Db 7126 ACCAGCTCAACCACTGCTGATCACTGCTGATCACTGATCACTGATCACTGATCACTGATCA 7185
Qy 2175 GlyPheLys1LeuArgTyArgSerAlaProTyCysSerLeuProArgAlaProLeuHisGly 2194
Db 7186 GGGCTTCAAGATCCGCTATTTCA----- 7206
Qy 2195 Phe1LeuGlnGlnThrSerThrGlnProGlyGlySer1LeuHisPheGlyCyAanAla 2214
Db 7206 ----- 7206
Qy 2215 GlyTyArgLeuVal1GlyHisSerMetAla1LeuCyThrArgHisProGlnGlyTyHis 2234
Db 7206 ----- 7206
Qy 2235 LeuTyPheSerGlyAla1LeuProLeuCyGlnAlaLeuSerCyGlyLupProGlnAlaPro 2254
Db 7207 -----GCCCTTCTCTGTGGGCTTCCCGAGGCCCCC 7236
Qy 2255 LybAang1LupMetVal1PheGlyLysGlyLupTyThrVal1GlyThrLysAlaVal1TySerCy 2274
Db 7237 AAGAAATGGAATGGGTGTGGCAAGGAGTACACAGTGGGACCAAGGCCATGTACAGCTGC 7296
Qy 2275 SerGlyGlyTyThr1eLeuGlnAlaGlyAaGlyAlaThrAlaGluCyValLeuAspThrGly 2294
Db 7296 ----- 7296
Db 7297 AGTAAAGCTTACCACTCCAGGCGAGGCTGAGCCACTGACAGATGTCTGACACAGGC 7356
Qy 2295 LeuTyPheAanAaAanVal1ProProGlnCyVal1ProVal1ThrCyProAspVal1Ser 2314
Db 7357 CTATGGAGCAACCCCAATGTCCACACAGTGTTCCTGTGACTTGTCTCGAATGTCACT 7416
Qy 2315 Ser1LeuSerVal1GlnHisGlyArgTyArgLeu1LeuPheGluThrGlnTyArg1PheGln 2334
Db 7417 AGCATACGCTGGAGCATGGCCGATGAGGCTTATCTTTAGACACAGTACATTCAG 7476
Qy 2335 AlaGlnLeuMetLeu1LeuCyAspProGly1TyTyTyTyThrGlyGlnArgVal1LeArg 2354
Db 7477 GCCCAGCTGATGCTCATCTGTGACCTTGCTACTACTGATGCTGATGCTGATGCTGATGCTG 7536
Qy 2355 CyGlnAlaAang1LupTyPheSerLeuGlyAspSerThrProThrCyVal1Le1LeSer 2374
Db 7537 TGTCAAGCCAAATGGCAATGGAGCTGGGAGCTTACGCGCCACCTCCGATCATCTCC 7596
Qy 2375 CyGlyGlyLupPro1LeuProProAang1HisArg1LeGlyThrLeuSerVal1TyGly 2394
Db 7597 TGTGGAAGCTCCCGATTTCCCAATGGCCACCGCATCGGACACTGTCTGTACGG 7656
Qy 2395 AlaThrAla1LeuPheSerCyAsnSerGlyTyThrLeuVal1GlySerArgVal1ArgGlu 2414
Db 7657 GCAACAGCATCTTCTCTGCAATTCGAGTACACACTGCTGGGCTCAGGGTGCCTGAG 7716
Qy 2415 CysMetAlaAang1LupLeuTyPheSerGlySerGlyVal1ArgCyVal1AlaGlyHisCyGly 2434
Db 7717 TGCATGGCCAAATGGGCTGTGGAGTGTGAAATCCGCTGCTGCTGCACTGGGG 7776
Qy 2435 ThrProGluPro1LeuVal1Aang1His1AlaAang1LupLupSerTyArgGlySer 2454
Db 7777 ACTCTGAGCCCATTTGCAACGACACATCAATGGGAGAACTACAGTACCGGGAGCT 7836
Qy 2455 Val1Val1TyGlnCyAanAlaGlyPheArgLeu1LeGlyMetSerVal1Arg1LeCyGln 2474
Db 7837 GTGGTGAACCAATGCAATGCTGGCTTCGCTGATGCGCATGTGTGGCATGTGCAG 7896
Qy 2475 GlnAspHis1eTyPheSerGlyLupTyThrProPheCyVal1Pro1LeuThrCyGlyHisPro 2494
Db 7897 CAGGATCATCACTGTGGGCAAGACCCCTTCTGTGTGCCAATTACTGTGGACACCA 7956
Qy 2495 GlyAanProVal1Aang1LupLeuThrGln1Val1Aang1PheAanLeuAanProVal1Lys 2514
Db 7957 GGCAACCTCTCAACGCTCTCACTCAGGGTACCAAGTTTAACCTCAACGATGTGTCAAG 8016
Qy 2515 PheVal1CyAanProGlyTyTyMetAlaGluGlyAlaAlaArgSerGlnCyVal1Aser 2534
Db 8017 TTTGTTTGCAACCTGGGTATATGCTGAGGGGCTGTAGTCCCATGCTGTGGCAGC 8076
Qy 2535 GlyGlnTyPheSerAspMetLeuProThrCyVal1Le1LeuAanCyThrAspProGlyHis 2554
Db 8077 GGGCAATGAGTGCAGAGCTGCCCACTGCAAGATCACTCACTGATCAAGATCCCTGGACAC 8136
Qy 2555 GlnGluAanSerVal1ArgGlnVal1HisAlaSerGlyProHisArgPheSerPheGlyThr 2574
Db 8137 CAAGAAATATGTGTTCTGTCAGTCCAGCCAGCGCGGCAAGETTCAGCTTCGGACCC 8196
Qy 2575 ThrValSerTyArgCyAanHisGlyPheTyTyLeuLeuGlyTyProVal1LeuSerCy 2594
Db 8197 ACTGTGTCTTAACGGGTCAACAGGCTTCTACTCTGGGCAACCCAGTGTCACTGC 8256
Qy 2595 GlnGlyAspGlyTyThrPheAspArgProArgProGlnCyValLeuVal1SerCyGlyHis 2614
Db 8257 CAGGAGATGAGCATGAGACCGTCCCGGCCCAAGTGTCTCTGGGTGTCTGTGGCAT 8316
Qy 2615 ProGlySerProProHisSerGlnMetSerGlyAspSerTyThrVal1GlyAlaVal1 2634
Db 8317 CCGGGCTCCCGCTCACTCCAGATGTCTGGAGACATTAATCTGGAGACATGTGCTG 8376
Qy 2635 ArgTySerCyVal1GlyLupArgTyThrLeuVal1GlyAanSerThrArgMetCyGlyLup 2654
Db 8377 CGGTACAGCTGCAATCGGCAAGCGTACTGTGTGGAAACAGACCCGCACTGTGGGCTG 8436
```

```

QY 2655 ASPGIYHISTRTPTHTGILSERLEUPROHISCYSESERGLYTHRSERYALGLVALCYSGLY 2674
DB 8437 GATGGACACTGGAGCTGGCTCCCTCCCTCACTGCTCAGGAACCAAGCGTGGAGCTTTGGCCGT 8496
QY 2675 ASPPROGLYILEPROLAHISGLYILEARGLEUGLYASPSETPHASPSETPROGLYTHRTVAI 2694
DB 8497 GACCCCTGGAGTCCCGGCTCATGGCATCCGTTTGGGGAGACAGTTTGAATCCAGGACCTGTG 8556
QY 2695 METARXPHSESECYSEGLIUALGLYHISVALILEUARGGLYSESESEGLUARGTHRCYSGIN 2714
DB 8557 ATGCGCTTCACTGCTGSAAGCTGACCAAGCTCCGAGGATCTCTCAGACGCAAGCTGTCTCA 8616
QY 2715 ALAASERGLYSETRIPSESEGLYSESEGLINPROGLUCYSGLYVALILESECYSEGLYASPRO 2734
DB 8617 GCCAATGCGCTCGTGAGCGCGCTCGCAGCCTGATGATGATGATCTTGTGGGAACCTC 8676
QY 2735 GLYTHTRPROSESEASPHALARGVALYALPHSESEAPSGLYLEUVALPHSESESESE1LE 2754
DB 8677 GGGACCTCCAAATGAATGCCGAGTGTTCAGTGAATGAGCTGTTCCTCAGCTCTATC 8736
QY 2755 VALTYRGLUCYARGLUGLYTYRTYRALATHRGLYLEUASERARGLHISCYSESEVAL 2774
DB 8737 GTCTATGATGTCGGGGAAGATACACCCACAGGCTGTCTCAGCCGCTCACTGCTCGCTC 8796
QY 2775 AENGLYTHRTPTHTGLYSEAPSPROGLUCYSEUVALILEANCYSGLYASPPROGLY 2794
DB 8797 AATGGTACCTGGACGAGCAGTGAACCTGAGTCCGTCGTCAATGAATCTGTGGTGAACCTCGG 8856
QY 2795 ILEPROALAENGLYLEUARGLEUGLYASPHASPHASPHASPHASPHASPHASPHASPHASPH 2814
DB 8857 ATTCAGGCAATGCGCTTCGCTGGGCAATACCTTCAGTACCAACAACTGTGACATAT 8916
QY 2815 GLNCYSEVALPROGLYTYRMECGLUSERHISARGVALSERVALILEUSECYSETHRTYR 2834
DB 8917 CAGTGTGCTCCCTGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8976
QY 2835 ASPARTHTRTPTAENGLYTHRTYRLEUPROVALCYSEVALILEUWECYSLYSPROPRO 2854
DB 8977 GACCGAGATGGAATGGAACCAAGCCGCTCGCAACCTCTCATGTGCAAGCAACCTCGC 9036
QY 2855 LEU1LEPROAENGLYLYSEVALYALYSEAPSPHEMETTRPGLYSESESEVAL1THRTYR 2874
DB 9037 CTCAATCCCAATGAGGAAGTGTGGGCTGTGACTTCATGTGGGCTCAAGTGTGACTTAT 9096
QY 2875 ALAcySELEUGLYTYRGLINLEUSESELEUPROVALAVALPHETHRCYSGLEUGLYAENGLY 2894
DB 9097 GCTGCTCTGAGAGGGGTACCAAGCTCTCCCTGCGCGGTTCACCTGTGAGGGAATGGG 9156
QY 2895 SETTRPTHTGLYGLUEUPROGLINCYSPHEPHEPROVALPHCYSEGLYASPPROGLYVALPRO 2914
DB 9157 TCTCGGACCGGAGAGCTGCTCATGTTTCTCTGTTCTGCGGGGATCTGTGTCCCG 9216
QY 2915 SEETHRGILYARGARGLEUAPARGLYPHSESETRYSERSESESESESESESESESESESESESE 2934
DB 9217 TCCCGTGGGAGGAGAGGACCGAGCTTCTCTCAAGCATCTGTCTCTCTCTCTCTCTCTCTCT 9276
QY 2935 HISPROPROLEUVALILEUVALYGLYSESEPROARGARGPHCYSEGLINSEAPSGLYTHRTYR 2954
DB 9277 CATCCCTCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9336
QY 2955 SESEGLYTHRGINPROSESECYSE1LEAPPROTHRTLEUTHRTCYAALASPPROGLYVAL 2974
DB 9337 AGTGGACCAAGCCAGCTGCAATATCCGACCTGACCAAGCTGTGCGGACCTGTGTGTGTGTGT 9396
QY 2975 PROGLINPHGLYILELINASPHASPHASPHASPHASPHASPHASPHASPHASPHASPHASPH 2994
DB 9397 CCACAGTTTGGAGTACGAACAATCTCAGGGCTCCAGGTGTGAAGACAGTCTCTCTCTCTCTCT 9456
QY 2995 ARGCYSEGLINLYSEGLYTYRLEUENGLINLYSETHRTTHRTTHRTCYSELEUPROASPHLEU 3014
DB 9457 CGTTGTCAAAAAGGCTACCTGCTTCAAGGCTCCACCAAGGACCTGCTCCCAAACTGTG 9516

```

---

```

QY 3015 THRTIPSESEGLYTHRTPROASPCYSEVALPROHISHCYSEARGLINPROGLINTHRTPRO 3034
DB 9517 ACCGGAGTGGAAACCCACCTGATGCTGTCCCAACCACTGAGGAGCCAGAGAGCA 9576
QY 3035 THHISALASPHANVALYALILEUAPLEUPROSESEMETGLYTYRTHRTLEU1LE-THRTPR 3054
DB 9577 ACCGATGCCAACCTCGGGGCTCGATTTGCCCTCATGGCTTACACGCTCATCTACTCC 9636
QY 3054 OALAARGTARGALASERPROSESEARVALA1APROSETHRALA1PROLA1AARGMECA1 3074
DB 9637 TGCCAGAGAGGCTTCTCTCTCAAGGCTGCTCGAGACCGGACCTGCAAGGCGGATGCG 9696
QY 3074 AALAGLYINALASERARGPROSESEALATTPARGSESEGLYPROVALGLYASPSESETH 3094
DB 9697 AGCTGAGACGAGCAAGCCGCTCATGCTGAGAGTCCGGCCCAATGGAGACCATCTAAC 9756
QY 3094 RLEUPROGLYSESEH1AARGSESEPROLYPRO 3104
DB 9757 ACTGCCGGAGGACCAAGCTCTCAACCAAGCTT 9787

```

RESULT 8  
ADH71145  
ID ADH71145 standard; DNA; 12900 BP.  
XX  
AC ADH71145;  
XX  
DT 25-MAR-2004 (first entry)  
XX  
DE Human gene of the invention NOV4f SEQ ID NO:41.  
XX  
KW de; gene; human; cytosolic; immunomodulator; neuroprotective; noctropic;  
KW anorectic; antidiabetic; antimicrobial; antileptic; gene therapy;  
KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;  
KW obesity; diabetes; infectious disease; metabolic syndrome X;  
XX  
XX  
OS Homo sapiens.  
XX  
PN W02003102155-A2.  
XX  
PD 11-DEC-2003.  
XX  
PF 03-JUN-2003; 2003WO-US017430.  
XX  
PR 03-JUN-2002; 2002US-0385120P.  
PR 04-JUN-2002; 2002US-0385784P.  
PR 05-JUN-2002; 2002US-0386041P.  
PR 05-JUN-2002; 2002US-0386047P.  
PR 06-JUN-2002; 2002US-0386376P.  
PR 06-JUN-2002; 2002US-0386453P.  
PR 06-JUN-2002; 2002US-0386664P.  
PR 06-JUN-2002; 2002US-0387016P.  
PR 07-JUN-2002; 2002US-0387966P.  
PR 07-JUN-2002; 2002US-038816P.  
PR 07-JUN-2002; 2002US-0386931P.  
PR 07-JUN-2002; 2002US-0386942P.  
PR 07-JUN-2002; 2002US-0386971P.  
PR 07-JUN-2002; 2002US-038762P.  
PR 08-JUN-2002; 2002US-0296960P.  
PR 10-JUN-2002; 2002US-0387400P.  
PR 10-JUN-2002; 2002US-0387535P.  
PR 11-JUN-2002; 2002US-0387610P.  
PR 11-JUN-2002; 2002US-0387625P.  
PR 11-JUN-2002; 2002US-0387634P.  
PR 11-JUN-2002; 2002US-0387668P.  
PR 11-JUN-2002; 2002US-0387696P.  
PR 11-JUN-2002; 2002US-0387702P.  
PR 11-JUN-2002; 2002US-0387836P.  
PR 11-JUN-2002; 2002US-0387859P.  
PR 12-JUN-2002; 2002US-0387933P.  
PR 12-JUN-2002; 2002US-0387934P.  
PR 12-JUN-2002; 2002US-0387960P.  
PR 12-JUN-2002; 2002US-0388022P.

|    |             |                 |
|----|-------------|-----------------|
| PR | 12-JUN-2002 | 2002US-0388096P |
| PR | 12-JUN-2002 | 2002US-0389123P |
| PR | 14-JUN-2002 | 2002US-0389118P |
| PR | 14-JUN-2002 | 2002US-0389120P |
| PR | 14-JUN-2002 | 2002US-0389144P |
| PR | 14-JUN-2002 | 2002US-0389146P |
| PR | 17-JUN-2002 | 2002US-0389729P |
| PR | 17-JUN-2002 | 2002US-0389742P |
| PR | 18-JUN-2002 | 2002US-0389884P |
| PR | 19-JUN-2002 | 2002US-0390066P |
| PR | 19-JUN-2002 | 2002US-0390209P |
| PR | 21-JUN-2002 | 2002US-0390763P |
| PR | 17-JUL-2002 | 2002US-0396706P |
| PR | 06-AUG-2002 | 2002US-0401628P |
| PR | 09-AUG-2002 | 2002US-0402156P |
| PR | 09-AUG-2002 | 2002US-0402256P |
| PR | 09-AUG-2002 | 2002US-0402389P |
| PR | 12-AUG-2002 | 2002US-0402786P |
| PR | 12-AUG-2002 | 2002US-0402816P |
| PR | 12-AUG-2002 | 2002US-0402821P |
| PR | 12-AUG-2002 | 2002US-0402832P |
| PR | 13-AUG-2002 | 2002US-0403448P |
| PR | 13-AUG-2002 | 2002US-0403459P |
| PR | 13-AUG-2002 | 2002US-0403531P |
| PR | 13-AUG-2002 | 2002US-0403532P |
| PR | 13-AUG-2002 | 2002US-0403563P |
| PR | 13-AUG-2002 | 2002US-0406317P |
| PR | 15-AUG-2002 | 2002US-0406317P |
| PR | 26-AUG-2002 | 2002US-0406182P |
| PR | 26-AUG-2002 | 2002US-0406355P |
| PR | 27-AUG-2002 | 2002US-0406240P |
| PR | 12-SEP-2002 | 2002US-0410084P |
| PR | 20-SEP-2002 | 2002US-0412528P |
| PR | 23-SEP-2002 | 2002US-0412731P |
| PR | 30-SEP-2002 | 2002US-0414801P |
| PR | 30-SEP-2002 | 2002US-0414839P |
| PR | 30-SEP-2002 | 2002US-0414840P |
| PR | 30-SEP-2002 | 2002US-0414954P |
| PR | 09-OCT-2002 | 2002US-0417186P |
| PR | 09-OCT-2002 | 2002US-0417406P |
| PR | 23-OCT-2002 | 2002US-0420639P |
| PR | 28-OCT-2002 | 2002US-0421156P |
| PR | 31-OCT-2002 | 2002US-0422690P |
| PR | 01-NOV-2002 | 2002US-0423130P |
| PR | 05-NOV-2002 | 2002US-0423798P |
| PR | 05-NOV-2002 | 2002US-0423798P |
| PR | 12-NOV-2002 | 2002US-0425453P |

```
Db 1417 TGCGTCTCTTCAG-----ACTGATGC---AGTGGCAGT---- 1449
Qy 257 ArglySerProlySerSerAsnAlaValGluLeuValAlaProGlyThrGluIleGlu 276
Db 1450 -----TCCCGGAGATTCAAGGCTTCTTATGAAGAGATCCAG 1485
Qy 277 GlnGlySerCysGlyAspProGlyIleProAlaTyrGlyArgArgGluGlySerArgPhe 296
Db 1486 CAGGCGAGTTGGGGGACCTTGACCTGATATGCGCGAGGAGGATCCCGGTTT 1545
Qy 297 HisHISGlyAspThrLeuLeuPheGlyCysGlnProAlaPheGluLeuValGlyGluPhe 316
Db 1546 CGCCACGGTGACACTCAAGTTTGAAGTCCAGCCCGCTTGAAGTGGTGAGACAGAG 1605
Qy 317 AlaIleThrCysGlnLeuAsnAsnGlnIlePheSerAlaValSerProGlyCysValPheSer 336
Db 1606 GCAATCATGATGCCAAGAAATACCAATGCTGGCTGAAGACCAAGCGCTGGTGTTC 1665
Qy 337 CysPhePheAsnPheThrSerProSerGlyValValLeuSerProAsnTyrProGluAsp 356
Db 1666 TGCTTCTTCAACTTCACAGCCCGTCTGGGGTGTGCTCTCCCACTACCGAGAGAC 1725
Qy 357 TyrGlyAsnHisIleuHisCysValTyrLeuIleLeuAlaArgProGluSerArgIleHis 376
Db 1726 TATGGCAACACCTTCACACTGTGTCTGCTCATCTGGCCAGGCTGAGAGCCGATCCAC 1785
Qy 377 LeuAlaPheAsnAspIleAspValGluProGlnPheAspPheLeuValIleValAspGly 396
Db 1786 CTGGCTTCAACGACATTGACGTGAGGCTCAGTTTATTTCCGTCATCAAGAGAGG 1845
Qy 397 AlaThrAlaGluAlaProValLeuGlyThrPheSerGlyAsnGlnLeuProSerSerIle 416
Db 1846 GCCACCGCGGGGCGCGCTCTGGGACCTTCTCAAGAAACGAGCTTCCCTCCATC 1905
Qy 417 ThrSerSerGlyHisValAlaArgLeuGluPheGlnThrAspHisSerThrGlyValArg 436
Db 1906 ACAAGCAGTGGCCACGTGGCCCGTCTGAGTTCCAGACTACACCTCCAGAGGAGAGG 1965
Qy 437 GlyPheAsnIleThrPheThrThrPheArgHisAsnGluCysProAspProGlyValPro 456
Db 1966 GGCTTCAACATCACTTTTACAGTAGTCTCCCAAGAGGCCGAGTCTGGCCCTTCCA 2025
Qy 457 ValAsnGlyIleValAspArgPheGlyAspSerLeuGlnLeuGlySerSerIleSerPheLeu 476
Db 2026 GTAAATGCAAAACGGTTTGGAGACAGCTCCAGCTGGGACGCTCCATCTCTCTCT 2085
Qy 477 AspGluGlyPheLeuGlyThrGlnGlySerGluThrIleThrCysValLeuLeuGluGly 496
Db 2086 GATGAAGGCTTCTTGGGACTCAGAGGCTCAGACATCACTGCTGCTGAGAGAGGCG 2145
Qy 497 SerValValTyrPheSerAlaValLeuArgCysGluAlaProCysGlyGlyHisIleuThr 516
Db 2146 AGGTGTCTGGAACACCGCTGTGCTGCGGTGAAGCTCCCTGTGTGTGCTGACT 2205
Qy 517 SerProSerGlyThrIleLeuSerProGlyTyrProGlyPheThrValAspAlaLeuSer 536
Db 2206 TCCCCACGCGCACATCTCTCTCCGGGCTGGCTGGCTTCAACAGAGATCCCTTGAC 2265
Qy 537 CysAlaTyrValIleGluAlaGlnProGlyTyrProIleValIleThrPheAspArgPhe 556
Db 2266 TGTGCTGGGTGTGATGAGGCCAGCGCATCCCATCAAAATCACTTCCACAGGTTT 2325
Qy 557 LysThrGluValAsnTyrAspThrLeuGluValArgAspGlyValArgThrTyrSerAlaPro 576
Db 2326 AAAACGAGGTCAACATGACACCCCTGGAAGTACGCGATGGCGGAGCTTACACGCGCC 2385
Qy 577 LeuIleGlyValTyrHisGlyThrGlnValProGlnPheLeuIleSerThrSerAsnTyr 596
Db 2386 TTGATGGGGTTTACACGAGGACCCAGGTTCCCGACTTCTCATGACCAACGACACTAC 2445
Qy 597 LeuTyrLeuLeuPheSerThrAspTyrSerHisSerAspIleGlyPheGlnLeuArgTyr 616
Db 2446 CTCTACTCTCTTCTTCTTACCAAGAGTCACTCGGACATGGGCTTCCAGCTCCGCTAT 2505
Qy 617 GluThrIleThrLeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGln 636
Db 2506 GAGGCTATAACACTGACAGTCAAGCACCTGTCTGGATCCAGAAATCCCATTAATGACAG 2565
Qy 637 ArgHisGlyAsnAspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyr 656
Db 2566 GGTCAATGGAATGACTTCACTGAGCGGCTGGAGCTTCACGCTGATCACTCGGCTAC 2625
Qy 657 ThrLeuSerAspGlyGluProLeuGluCysGluProAsnPheGlnIlePheSerArgAlaLeu 676
Db 2626 ACATTAAAGTGAAGGAGGAGCTTGAAGTGTGAGCCCACTTCAATGAGCGGCGCTG 2685
Qy 677 ProSerCysGluAlaLeuCysGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSer 696
Db 2686 CCCAGTTTGAAAGCTCTCTGTGGCTTCAATTAAGGCTCAAGTGGAGCATCTTGTG 2745
Qy 697 ProGlyPheProAspPheTyrProAsnLeuAsnGlyThrTyrIleGluThrSer 716
Db 2746 CAGGTTTCCCTGACTTCTTACCCCAACACTTGAACCTGCACTGATTAATCAACATCT 2805
Qy 717 HisGlyValGlyValPhePheThrPheHisIleGluLeuSerGlyHisAspTyr 736
Db 2806 CATGGCAAGGAGTGTCTTCACTTCCACACTTCCAGTGAAGTGGCATGACTAC 2865
Qy 737 LeuLeuIleThrGluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArg 756
Db 2866 CTCTCATCACTGAAGAGGAGGAGCTTCAACGAGCCCTGAGAGGAGCTTAAGTGAATCTCGG 2925
Qy 757 LeuProAlaProIleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIle 776
Db 2926 CTGCAAGCTCCCATAGCGCTGGGCTCTATGGCACTTCACTGAGGAGTGGCTTCACTC 2985
Qy 777 SerAspPheSerMetSerTyrGluGlyPheAsnIleThrPheSerGluTyrAspLeuGlu 796
Db 2986 TCTGATTTCTCATGTATGAAAGATTCAACATCACTTCTCAGAGTACGACTGGAG 3045
Qy 797 ProCysGluGluProGluValProAlaTyrSerIleArgGlyGlyLeuGlnPheGlyVal 816
Db 3046 CCTGTGAGAGCCGAGAGTCCAGCTCAAGCATCCGGAAGGCGCTTGCAGTTTGGCGTG 3105
Qy 817 GlyAspThrLeuThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIle 836
Db 3106 GCGGACACTTGACCTTCTCTGCTTCCCGGATACCGTCTGAGAGGACCGCCGCACTC 3165
Qy 837 ThrCysLeuGlyGlyValArgArgLeuTyrPheSerProLeuProArgCysValAlaGlu 856
Db 3166 ACGTGTGGGGGACAGACGCGCTGTGAGGCTCGCTTGCACAGAGTGTGTGTGAG 3225
Qy 857 CysGlyAsnSerValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsn 876
Db 3226 TGTGGAAITTCAGTCAAGGACACTCAAGGTAATTTGCTGTCCCAACTTTCGTGTAAC 3285
Qy 877 TyrAsnAsnAsnHisGlyCysIleTyrSerIleGlnThrGlnProGlyValGlyIleGln 896
Db 3286 TACAATAACATCATGATCATCTCACTCATCAAGCCAGCAGGAGGAGGAAATTCAG 3345
Qy 897 LeuLysAlaArgAlaPheGluLeuSerGluGlyAspValLeuLysValTyrAspGlyAsn 916
Db 3346 CTGAAGCCAGGGGATTCGAACCTTCGAAGAGAGTCTCTCAAGGTTTATGATGGCAAC 3405
Qy 917 AsnAsnSerAlaArgLeuLeuGlyValPheSerHisSerGluLeuThrGlyValThrLeu 936
Db 3406 AACAACTCCCGCGTGTGCTGGAGTTTAAAGCATTTTGAGATGATGGGGAGACTTTG 3465
Qy 937 AsnSerThrSerSerSerLeuTyrPheAspPheIleThrAspAlaGluAsnThrSerLys 956
Db 3466 AACAGCAGTCCAGAGCTGTGTGTGATTTGATCATCATGATCTGAATAACACACCAAG 3525
Qy 957 GlyPheGluLeuHisPheSerSerPheGluLeuIleValCysGluAspProGlyTyrThrPro 976
Db 3526 GGCTTGAATCGACTTTTCCAGTTTGAATCTATCAAAATGTGAGAGCCACGAGAACCCCC 3585
```

```
Qy 977 LysPheGlyTyrLeuValHisAspGluGlyHisPheAlaGlySerSerValSerPheSer 996
Db 3586 AAGTTGGCTACAAAGCTTCAATGATGAGCTATTTGAGGAGACTCCGTGTCTTCCACG 3645
Qy 997 CysAspProGlyTyrSerLeuArgLysSerGluLeuLeuCysLeuSerGlyGluArg 1016
Db 3646 TGTGACCTCGGATACAGCTCGGGGGTAGTGAAGAGCTGTGTCTGATGAGAGAGCG 3705
Qy 1017 ArgThrTrpAspArgProLeuProThrCysValAlaGluCysGlyGlyThrValArgGly 1036
Db 3706 CGGACCTGGAGCCGGCTCTGCCACCTGTGCGCGAGTGGAGGAGCAGTGAAGGA 3765
Qy 1037 GluValSerGlyGlnValLeuSerProGlyTyrProAlaProTyrGlnHisAsnLeuAsn 1056
Db 3766 GAGGTGTGGGGGAGGTCTGTCTACCCGGGTATCCAGCTCCCTTTGAAACAATCTCAAC 3825
Qy 1057 CysIleThrThrIleGluAlaGlyCysThrIleGlyLeuHisPheLeuValPhe 1076
Db 3826 TGCATCTGGACCATCGAAGCAGAGCGGGCTGCACCATTTGGGCTACACTTCTGTGTTT 3885
Qy 1077 AspThrGluGluValHisAspValLeuArgIleThrAspGlyProValGluSerGlyVal 1096
Db 3886 GACACAGAGAGGTTCCAGACGTGCTGCGCATCTGGGATGGGCGCTGTGAGAGCGGGGTT 3945
Qy 1097 LeuLeuLysGluLeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSer 1116
Db 3946 CTGCTGAAGAGCTGAGTGGCCGGGCTCTGCCCAAGACCTGCAATGACCTTCAACTCG 4005
Qy 1117 ValValLeuGlnPheSerThrAspPheThrSerLysGlnGlyPheAlaIleGlnPhe 1136
Db 4006 GTGCGCTCGAGCTCAGCATGACTTCTTCCACGAAAGCAGGGGCTTTGCCATTCATTT 4065
Qy 1137 SerValSerThrAlaThrSerCysAsnAspProGlyIleProGlnAsnGlySerArgSer 1156
Db 4066 TCAGGTTCCACAGAACCTCTCTGCAATGACCTGGGATCCCGCAGAAATGGGAGTCGGAGT 4125
Qy 1157 GlyAspSerTrpGluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeu 1176
Db 4126 GGTACACTGGGAGGCGGGGACTCCACAGGTTCACAGTGAACCTCGGTACAGCGGTG 4185
Qy 1177 GlnIleSerAlaGluIleSerCysValLysIleGluAsnArgPhePheTrpGlnProSer 1196
Db 4186 CAGGAGAGTCAGAGATCAGTGTGTGAAGATCGAAGACAGTCTTCTGGCACCCAGC 4245
Qy 1197 ProProThrCysIleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeu 1216
Db 4246 CCGGCAACATGATCGCTCCCTGGCGGGGAGACCTGACAGGACCATCTGGAGTCATCTC 4305
Qy 1217 SerProAsnTyrProGluProTyrProProGlyLysGluCysAspTrpLysValThrVal 1236
Db 4306 TCACCAAAATTAACCAAGAACCTTACCCGCAAGGAGAGTGTGATGGAAAGTACCGTC 4365
Qy 1237 SerProAspTyrValIleAlaLeuVal---PheAsnIlePheAsnLeuGluProGlyTyr 1255
Db 4366 TCACCAAGACTACGTCAATCGCCCTGTGATGTTCCTCCACGCTTAACTGGAGCCTGGCTAT 4425
Qy 1256 AspPheLeuHisIleTyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyr 1275
Db 4426 GACTTCCCTCCATCTACGAGGAGCGGACTCTCTCAGCCCTCATAGGAAGCTTCTAT 4485
Qy 1276 GlySerGlnLeuProGlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArg 1295
Db 4486 GGGCTCCAGCTCCAGGGCGGATTTGAAGACAGCAACACGCTTCTCGCTTCGCGC 4545
Qy 1296 SerAspAlaSerValSerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGlu 1315
Db 4546 AGCATGTGATCTGTGAAGCAATGCGTCTGTCTGATTCCTCCAGAAAAACCCGGGAG 4605
Qy 1316 SerCysPheAspProGlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeu 1335
Db 4606 TCATGTTTGTATCTGTGTTTCATCAAGAACGCGACACGGGTGGGGTCCGACTGAACCTG 4665

Qy 1336 GlySerSerValThrTyrTyrCysHisGlyGlyTyrGluValGluGlyThrSerThrLeu 1355
Db 4666 GGGCTCTCCGTCACTCACTCACTGCGCACGGGGGCTACGAAATTTAGGGCACCTGACCTG 4725
Qy 1356 SerCysIleLeuGlyProAspGlyLysProValIleTrpAsnAspProArgProValCysThr 1375
Db 4726 AGTGCATCTCGGGGGCTGATGGGAAGCCGTTGGAAACAATCCCCGGCCAGTGTGCACA 4785
Qy 1376 AlaProCysGlyGlnGlyThrValGlySerAspGlyValValLeuSerProAsnTyrPro 1395
Db 4786 GCGCCCTGTGGGGGACAGTATGTGGGTGTGGAGCGAATGTGTTCTCCCACTACCC 4845
Qy 1396 GlnAsnTyrThrSerGlyGlnIleCysLeuTyrPheValThrValProLysAspTyr--- 1414
Db 4846 CAGAACTACACCAATGAGACAGATCTGTGATTTTGTACTGTGGCCCAAGACTAGT 4905
Qy 1415 ValValPheGlyGlnPheAlaPhePheHisThrAlaLeuAsnAspValAlaGluValHis 1434
Db 4906 GTGGTGTGGCCAGTTCGCTTCCTTTCACACGGCCCTCAACGACGTGTGAGGTTCCAC 4965
Qy 1435 AspGlyHisSerGlnHisSerArgLeuLeuLeuSerSerLeuSerGlySerHisThr---Gly 1453
Db 4966 GACGGCCACAGCCAGCAGCTGCGGCTCTCAGCTCCCTCGGGCTCCCATACAGGTGA 5025
Qy 1454 GluSerLeuProLeuAlaThrSerAsnGlnValLeuIleLysPheSerAlaLysGlyLeu 1473
Db 5026 GAATCACTGCCCTTGGCCACCTCCATCAAGTTCTCTTAAGTTACAGCCCAAAAGCCTC 5085
Qy 1474 AlaProAlaArgGlyPheHisPheValTyrGlnAlaValProArgThrSerAlaThrGln 1493
Db 5086 GCACACAGCAGAGGCTTCCACTTGTGTACCAAGCGGTTCTCGAACACGCGCACGACAG 5145
Qy 1494 CysSerSerValProGluProArgTyrGlyLysArgLeuGlySerAspPheSerValGly 1513
Db 5146 TGCAGCTCTGTGCGGAACCCGCTATGGCAAGAGCGTGGCAGTGACTTCTCGTGGGG 5205
Qy 1514 AlaIleValArgPheGluCysAsnSerGlyTyrAlaLeuGlnGlySerProGluIleGlu 1533
Db 5206 GCCATCGTCCGCTTCAATCAACTCCGCTTAGCCCTGACGAGGGGTGCCAAGATGAG 5265
Qy 1534 CysLeuProValProGlyAlaLeuAlaGlnTrpAsnValSerAlaProThrCysValVal 1553
Db 5266 TGCTCCCTGCTGCTGGGGCTTGGCCCAATGGAATGTCACAGGCCACAGTGTGTGCTG 5325
Qy 1554 ProCysGlyGlyAsnLeuThrGluArgArgGlyThrIleLeuSerProGlyPheProGlu 1573
Db 5326 CCGGTGTGAAGCACTCAACAGAGCGGAGGCCATCTCTGCTCCCTGCTCCAGAG 5385
Qy 1574 ProTyrLeuAsnSerLeuAsnCysValTrpLysIleValValProGluGluAlaGlyIle 1593
Db 5386 CCGTACTTCAACAGCTCAACTGTGTGTGGAAGATCGTGGTCCCGAAGGCGCTGGCATC 5445
Qy 1594 GlnIleGlnValValSerPheValThrGluGlnAsnTrpAspSerLeuGluValPheAsp 1613
Db 5446 CAGATCCAAATTCAGATTGTGTGACAGACAGAACTGGGACTCGGTGAAGATTTGAT 5505
Qy 1614 GlyAlaAspAsnThrValThrMetLeuGlySerPheSerGlyThrThrValProAlaLeu 1633
Db 5506 GGTGAGATTAACCTGAACCATAGCTGGGAGTTTCCAGGAACAACGTCGCTGCCCTT 5565
Qy 1634 LeuAsnSerThrSerAsnGlnLeuTyrLeuHisPheTyrSerAspIleSerValSerAla 1653
Db 5566 CTGAACGACCTCCAAACGACTCACTTCAATTTCACTCAAGATATACAGCGATTCGCA 5625
Qy 1654 AlaGlyPheHisLeuGluTyrTyrThrValAlaGlyLeuSerSerCysProGluProAlaVal 1673
Db 5626 GCTGGCTTCCACTTGGAGTACAAACGATGGGCTGAGCAGTGTCCGAAACCTGGCTGTG 5685
Qy 1674 ProSerAsnGlyValLysThrGlyGluArgTyrLeuValAsnAspValValSerPheGln 1693
Db 5686 CCCAGTAAACGGGGTGAAGACTGGCAGCGCTACTTGTGTAATGATGTGGTGTCTTTCAG 5745
Qy 1694 CysGluProGlyTyrAlaLeuGlnGlyHisAlaHisIleSerCysMetProGlyThrVal 1713
```

```
Db 5746 TGTGAGCCGGATATGTCCTCCAGGGCCAGCCCACTCTGCAATGCCGAACAGTGT 5805
Qy 1714 ATGAGTTPAANTYRProProProLeuCySIIlealagInCySgIyThrValGluGlu 1733
Db 5806 CGGCATGTGAACTTACCTCTCTCCACTGTATTTGACAGTGTGGGGAACAAGTGAAGAG 5865
Qy 1734 MetGluGluValIleLeuSerProGlyPheProGlyAsnTYRProSerAsnMetCys 1753
Db 5866 ATGGAGGGGGTATCTTGAAGCCCCGGCTTCCAGGCAACTACCCAGTAACATGAGATGC 5925
Qy 1754 SerTPlySIIleIleuProValGlyPheGlyAlaIleGlnPheLeuAsnPheSer 1773
Db 5926 TCCTGAAAATAGCACTGCCGCTGGGCTTTGGAGCTCAATCCAGTTCTTAATCTTCC 5985
Qy 1774 ThrGluProAsnHisAspTYRIleGluIleArgAsnGlyProTYRGLuThrSerArgMet 1793
Db 5986 ACCGAGCCCAACCAAGACTTACATAGAAATCCGGAATGGCCCCCTATGAGACCAAGCCGATG 6045
Qy 1794 MetGlyArgPheSerGlySerGluLeuProSerSerLeuLeuSerThrSerHisGluThr 1813
Db 6046 ATGGAGAAATTCAGTGAAGGAGCTTCCAGCTCCCTCTCTCCAGCTCCACAGAC 6105
Qy 1814 ThrValTYRphenIleSerAspHisSerGlnAsnArgProGlyPheGlySLeuGluTYRGLn 1833
Db 6106 ACCGTGATTTTCCACAGGACCACTCCAGAAATCGGCCAGGATTCAAGCTGGAGTATCAG 6165
Qy 1834 AlaTYRGLuLeuGlnGluCysProAspProGluProPheAlaAsnGlyIleValArgGly 1853
Db 6166 GCCTATGAACTTCAAGAGTCCCAAGCCCAAGAGCCCTTGGCAATGGCAATGGAGAGGA 6225
Qy 1854 AlaGlyTYRAsnValGlyGlnSerValThrPheGlnCysLeuProGlyTYRGLuLeuThr 1873
Db 6226 GCTGCTACAAAGTGGGACATCAGTGACCTTGAGTGGCTCCGGGGTATCAATTGACT 6285
Qy 1874 GlyHisProValLeuThrCysGlnHisGlyThrAsnArgAsnTPRAspHisProLeuPro 1893
Db 6286 GGCAACCTGTCTCAAGGTGCATGACGACCAAGCCGGAATCGGACCAACCCCTGGCC 6345
Qy 1894 LysCysGluValProCysGlyGlyAsnIleThrSerSerAsnGlyThrValTYRserPro 1913
Db 6346 AAGGTGAAAGTCCCTTGGGGGGAACATCATCTTCCACAGGCAT -GTGATCTCCGG 6404
Qy 1914 GlyPheProSerProTYRserSerSerSerGlnAspCysValTPRLeuIleThrValProIle 1933
Db 6405 GGGTTCCTTACCCCGTACTCCAGCTCCAGACTGTGTCTGGCTGATCACCGTGCATTC 6464
Qy 1934 GlyHisGlyValArgLeuAsnLeuSerLeuLeuGlnThrGluProSerGlyAspPheIle 1953
Db 6465 GGCAATGGCTCCGCTCAACCTCAAGCTCTGCAAGACAGCCCTTGAGATTTTCATC 6524
Qy 1954 ThrIleTPRAspGlyProGlnGlnThrAlaProArgLeuGlyValPheThrArgSerMet 1973
Db 6525 ACCATCTGGAGTGGGCACAGCAACAGCACAGCGCTGGGCTTCCACCCGAGACAG 6584
Qy 1974 AlaIleValLeuThrValGlnSerSerSerAsnGlnValLeuLeuLysPheHisArgAspAla 1993
Db 6585 GCCAAGAAACAAGTGAAGTTCATCAACAGAGTCTCTGCAAGTTCCACCGTATGCA 6644
Qy 1994 AlaThrGlyGlyIlePheAlaIleAlaPheSerAlaTYRProLeuThrLysCysProPro 2013
Db 6645 GCCACAGGGGGGAGTCTTCCGCAATGCTTCTCCGCTTATCCATCCACCAATGCCCCCTCT 6704
Qy 2014 ProThrIleLeuProAsnAlaGluValValThrGluAsnGluGlnPheAsnIleGlyAsp 2033
Db 6705 CCCACATCTCCCAACGCCGAAAGTGTACACAGAAATGAAGAAATTCATTAAGTATC 6764
Qy 2034 IleValArgTYRArgCysLeuProGlyPheThrLeuValGlyAsnGluIleLeuThrCys 2053
Db 6765 ATGTATCGTACAGATGCTCTCTGCTTACCTTATGAGTGGGAAATTCATTCAGCTGC 6824
Qy 2054 LysLeuGlyThrTYRLeuGlnPheGluGlyProProProIleCysGluValHisCysPro 2073
Db 6825 AAACTTGAACCTTACCTGAGATTGAAAGCAACCCCGATATGTGAAGTGCATGTCCA 6884
Qy 2074 ThrAsnGluLeuLeuThrAspSerThrGlyValIleLeuSerGlnSerTYRProGlySer 2093
Db 6885 ACAAATGAGCTTTGACAACTTCCACAGGCGAGATCTTAAGCAGAGCTTACCTGGAAC 6944
Qy 2094 TYRProGlnPheGlnThrCysSerTPRLeuValArgValGluProAspTYRAsnIleSer 2113
Db 6945 TATCCCAATTCAGACCTGTCTTGGCTGTGAGAGTGAAGCCGACATATAACATCTTC 7004
Qy 2114 LeuThrValGluTYRphenSerGluGlyArgGlnTYRAspGluPheGluIlePheAspGly 2133
Db 7005 CTCACAGTGAAGTCTCTCAAGCAAGAAATGAAATGAAATTTGAGATTTTATGATGT 7064
Qy 2134 ProSerGlyGlnSerProLeuLeuLysAlaLeuSerGlyAsnTYRSerAlaProLeuIle 2153
Db 7065 CCATCAGACAGAGTCTCTGCTGAAGCCCTCAGTGGAAATTAATCTCAAGTCCCTGATT 7124
Qy 2154 ValThrSerSerSerAsnSerValTYRLeuArgTPRSerSerAspHisAlaTYRAsnArg 2173
Db 7125 GTCCACAGCTCAAGCAACTGTGTACTGCTGTGATCATCTGATCACGCCCTACAAATCG 7184
Qy 2174 LysGlyPheIleValArgTYRserAlaProTYRserSerLeuProArgAlaProLeuHis 2193
Db 7185 AAGGCTTCAGATCCGCTATTTCAAGCCCTTATCTGACACCTGGCCAGGGCTTCATCTCAT 7244
Qy 2194 GlyPheIleLeuGlyGlnThrSerThrGlnProGlyGlySerIleHisPheGlyCysAsn 2213
Db 7245 GGCCTCATCTTACGACAGACAGCAACCCAGCCGGGCTCATCTCACTTTGGCTGCAC 7304
Qy 2214 AlaGlyTYRArgLeuValGlyHisSerMetAlaIleCysThrArgHisAspProGlnGlyTYR 2233
Db 7305 GCCGCTTACCGCTGTGTGGACACAGCAAGTGGCACTGTGTACCCGACACCCCAAGGGCTAC 7364
Qy 2234 HisLeuTPRserGluAlaIleProLeuCysGlnAlaLeuSerCysGlyLeuProGluAla 2253
Db 7365 CACTGTGAGAGGAAGCATCTCTGTGTAACTCTTCTTGGCTTCTCGAAGGCC 7424
Qy 2254 ProIysAsnGlyMetValPheGlyLysGluTYRThrValGlyThrLysAlaValTYRser 2273
Db 7425 CCCAAGATGGAAGTGTGTTGGCAAGAGTACACAGTGGGAAACCAAGGCCATGTACAGC 7484
Qy 2274 CysSerGluGlyTYRHisSLeuGlnAlaGlyAlaIleThrAlaIleGluCysLeuAspThr 2293
Db 7485 TGCAGTGAAGCTACACCTTCAAGCGAGCTAGAGCCATGTGAGAGTGTCTGGACACA 7544
Qy 2294 GlyLeuTPRserAsnArgAsnValProProGlnCysValProValThrCysProAspVal 2313
Db 7545 GGCTTATGAGCAACCCCAATGTCTCCACACAGTGTCTCCCTGTGATTTGTCTGATGTC 7604
Qy 2314 SerSerIleSerValGluHisGlyArgTPRArgLeuIlePheGluThrGlnTYRGLnPhe 2333
Db 7605 AGTACATCAAGCGTGGAGCAAGGCCGAAAGGCTTATCTTTAGACACAGTATCACTTC 7664
Qy 2334 GlnAlaGlnLeuMetLeuIleCysAspProGlyTYRTYRTYRThrGlyGlnArgValIle 2353
Db 7665 CAGGCCAGCTGAGTCACTGATGACCTGGCTACTACTATCTGCGCAAAAGGCTATC 7724
Qy 2354 ArgCysGlnAlaAsnGlyLysTPRserLeuGlyAspSerThrProThrCysArgIleIle 2373
Db 7725 CGCTGTAGGCAATGGAATGAGAGCTCTGGGGAATCTTACGCCCACTTCCGAATCATC 7784
Qy 2374 SerCysGlyGluLeuProIleProProAsnGlyHisArgIleGlyThrLeuSerValTYR 2393
Db 7785 TCCTGTGGAAGCTCCCGATTTCCCCCAATGAGCCACGCAATCGGAACACTGTGTGCTAC 7844
Qy 2394 GlyAlaThrAlaIlePheSerCysAsnSerGlyTYRThrLeuValGlySerArgValArg 2413
Db 7845 GGGGCAACAGCAATCTTCTCTCAATTCGGAATACACTGTGTGGCTCCAGAGGTGCGT 7904
Qy 2414 GluCysMetAlaAsnGlyLeuTPRserGlySerGluValArgCysLeuAlaGlyHisCys 2433
Db 7905 GAGTGAAGGCAATGAGCTGTGAGTGGCTGTGAAGTCCGCTTCTGTGACACTGT 7964
```

|    |      |  |      |
|----|------|--|------|
| QY | 2434 | LIYTHPRROGLIYPRQIIYVAlAsnGIYHIIeAEnGIYGIuAsnTYSerTYAArgIY        | 2453 |
| Db | 7965 | GGACCTCTGGACCCCATTTGTCAAGGACACATCAATGGAGAACTAACAGCTACCGGGGC      | 8024 |
| QY | 2454 | SeValValIYrGIInCyAsnAlaGIYPhEAArgLeuIleGIYMeSerValArgIleCyS      | 2473 |
| Db | 8025 | AGTGGGTGTACCAATGGCATGCTGGCTTCGGCCGATCGGAGATGCTGTGGCATCTGC        | 8084 |
| QY | 2474 | GIInGIaAPRHIIeITPSeGIYLYsThrPRORHeCyValProIleThCySeGIYHIS        | 2493 |
| Db | 8085 | CAGCGAGATCATCACTGGTCGGGCAAGACCCCTTCTGTGTGTG-----                 | 8126 |
| QY | 2494 | ProGIYAnProValAsnGIYLeuThrGIInGIYAsnGIInPhEAuLeuAsnAlaValVal     | 2513 |
| Db | 8126 | -----  | 8126 |
| QY | 2514 | LYsPheValCyAsnProGIYTYrMeAlaGIuAlaAlaArgSeGIInCyLeuAla           | 2533 |
| Db | 8126 | -----  | 8126 |
| QY | 2534 | SeGIYGIInTPSeAPMeLeuProThrCyAsnIleIleAsnCyThAPProGIY             | 2553 |
| Db | 8126 | -----  | 8126 |
| QY | 2554 | HIeGIInGIuAsnSeValArgGIInValHISAlaSeGIYPROHISeAPhSeAPheGIY       | 2573 |
| Db | 8126 | -----  | 8126 |
| QY | 2574 | ThrThrValSeTYrArgCyAsnHISGLYrPheTYrLeuLeuGIYThrProValLeuSer      | 2593 |
| Db | 8126 | -----  | 8126 |
| QY | 2594 | CySeGIInGIYAsnGIYThrTrAPArArgProArgProGInCySLeuLeuValSeCYSeGIY   | 2613 |
| Db | 8127 | -----CTGGTGTCTGTGGC  | 8141 |
| QY | 2614 | HIePRROGIYSerProProHISeGIInMeSeGIYAsnSerTYrThrValGIYAlaVal       | 2633 |
| Db | 8142 | CATCCGGGCTCCCCCTCACTCCAGATGTGTGGAGACAGTAACTGTGGAGACATGG          | 8201 |
| QY | 2634 | ValArgTYrSerCySIIeGIYLYsArgThrLeuValGIYAsnSerThAPArgMeCYSeGIY    | 2653 |
| Db | 8202 | GTGGCGATCAGTGTGATGGGCAAGGATCTGTGGTGGAAACAGACCCGATGTGTGG          | 8261 |
| QY | 2654 | LeuAPRGIIHISTrPThrGIYSeIeLeuProHISCySeSeGIYThrSeValGIYValCyS     | 2673 |
| Db | 8262 | CTGGATGGACCTGGACCTGGCTCCCTCCCTCACTCTCAAGAACAGGCTGGAGATTGGC       | 8321 |
| QY | 2674 | GIYAsnPRROGIYIleProAlaHISGLYIleArgLeuGIYAsnSerPhAsnPRROGIYThr    | 2693 |
| Db | 8322 | GGTGCACCTGGGATCCCGGCTCATGGCATCCGTTTGGGGAGACAGCTTGTGATCCAGGCACT   | 8381 |
| QY | 2694 | ValMeArGPheSeSeCYGIuAlaGIYHISValIleuArGGIYSerSeGIYAlaThrCyS      | 2713 |
| Db | 8382 | GTGAAGCCCTTCACTGTGAAGCTGGCAAGCTGCTCCGGGATGCTCAAGACGCACTGT        | 8441 |
| QY | 2714 | GIuAlaAsnGIYSerTrPSeSeGIYSeGIInProGIYSeGIYValIISeSeCYSeGIYAsn    | 2733 |
| Db | 8442 | CAAGCCAAATGGCTCTGTGAGCGGCTCGCAGCCTGAGTGTGAGATGATCTTTTGGGAAC      | 8501 |
| QY | 2734 | ProGIYThrPRORSeAPsnAlaArgValValPheSerAPRGIIYLeuValPheSeSeSer     | 2753 |
| Db | 8502 | CTTGGGACCTCCAAAGTAATGCCGAAATGTGTCAATGATATGGCTGTGTTTCTCCAGCTCT    | 8561 |
| QY | 2754 | IIeValIYrGIInCyAsnGIuGIYTYrTYrAlaThrGIYLeuLeuSeSeArgHISCySeSer   | 2773 |
| Db | 8562 | ATCGCTATGAGTGGCGGGAGAGAACTATACGCCACAGGCGTGTACGCGTCACTGCTCG       | 8621 |
| QY | 2774 | ValAsnGIYThrTrPThrGIYSeArAsnPRROGIYCySLeuValIISeAsnCySeGIYAsnPro | 2793 |
| Db | 8622 | GTCAATGTATCTGGACAGGACGTGACCTGTAGTGTCTTCGTATTAATCTGTGGTGAACCT     | 8681 |

|          |  |  |      |
|----------|--|--|------|
| QY       | 2794   | GlyIleProAlaAsnGlyLeuAArgLeuGlyValAsnAspPheAArgTyrAsnLeuThrValThr  | 2813 |
| Db       | 8662   | GGAATTCAGCCAAATGGCTTCGGCTGGGCATATGACTTCAGAGTCAACAAACTGTGACA        | 8741 |
| QY       | 2814   | TyrGlnCysValProGlyTyrMetMetGluSerHisArgValSerValLeuSerCysThr       | 2833 |
| Db       | 8742   | TATCATGATGTGCTCCTGGCTATATATGAGTCAATATGAGTATCTGTGCTGAGCTGACCC       | 8801 |
| QY       | 2834   | LysAspArgThrTyrPAsnGlyThrLysProValCysIlyValAlaLeuMetCysLysProPro   | 2853 |
| Db       | 8802   | AAGGACCGGACATGGAATGGAAACCAAGCCCTGTGCAGAAAGCTCATGTGCACACCACT        | 8861 |
| QY       | 2854   | ProIleuLeProAsnGlyLysValValGlySerAspPheMetTrpGlySerSerValThr       | 2873 |
| Db       | 8862   | CCGCTCATCCCAATGGGAAGGTGTGGGGTCTAGCTTCATGTGGGCTCAAGTGTACT           | 8921 |
| QY       | 2874   | TyrAlaCysLeuGlnGlyTyrGlnLeuSerLeuProAlaValPheThrCysGlnGlyAsn       | 2893 |
| Db       | 8922   | TATGCTGTGCTGGAGGGGTACCACTCTCCTGCCCCGGGTGTTCACTGTGAGGGAAT           | 8981 |
| QY       | 2894   | GlySerTrpThrGlyGluLeuProGlnCysPheProValPheCysGlyAspProGlyVal       | 2913 |
| Db       | 8982   | GGGTCTGTGACCGGAGAGACTCTCAGTGTTCCTGTGTTCTCGGGGGAATCTGTGTCTC         | 9041 |
| QY       | 2914   | ProSerArgGlyValArgArgGlyLysAspArgLysPheSerTyrArgSerSerValSerPheSer | 2933 |
| Db       | 9042   | CCGTCCCGCTGGGAAGAGAGAGACCGAGGCTTCTCCACAGTATCTGTCTCTTCTCC           | 9101 |
| QY       | 2934   | CysHisProProLeuValLeuValGlySerProArgArgPheCysGlnSerAspGlyThr       | 2953 |
| Db       | 9102   | TGCAATCCCCCTCTGTGTCTGTGTGGGCTCTCCAGCAGGTTTTCAGTCAATGGAGCA          | 9161 |
| QY       | 2954   | TrpSerGlyThrGlnProSerCysAlaAspProThrLeuThrThrCysAlaAspProGly       | 2973 |
| Db       | 9162   | TGAGGTGGCACCGACCGACGTGATGATCCAGCCGTGACAGCACTGTGGCGGACCTGTGT        | 9221 |
| QY       | 2974   | ValProGlnPheGlyIleGlnAsnAsnSerGlnGlyTyrGlnValGlySerThrValLeu       | 2993 |
| Db       | 9222   | GTGCCACAGTTTGGGAATACAGAACAAATCTCAGGGCTACCAAGTTTGAAGCACAGTCTCTC     | 9281 |
| QY       | 2994   | PheArgCysGlnLysGlyTyrLeuLeuGlnGlySerThrThrArgThrCysLeuProAsn       | 3013 |
| Db       | 9282   | TTCGGTGTCAAAAGGCTACCTCTTCAGGGCTCCACACACAGAGACTGTCTCCAAAC           | 9341 |
| QY       | 3014   | LeuThrTrpSerGlyThrProProAspCysValProHisHisCysArgGlnProGluThr       | 3033 |
| Db       | 9342   | CTGACCTGGAGTGAACCCCACTCGACTGTGTCCCCACCACTGACAGCAGCCAGAGAGCG        | 9401 |
| QY       | 3034   | ProThrHisAlaAsnValGlyValAlaLeuAspLeuProSerMetGlyTyrThrLeuIle-Th    | 3053 |
| Db       | 9402   | CCAACGCAATGCCAAGTCGGGGCCCTGGATTTCGCCCTCAATGGGCTTCACGCTCATCTAC      | 9461 |
| QY       | 3053   | TrpAlaLysArgArgLysSerProSerArgValAlaProSerThrLysProAlaAspArgMe     | 3073 |
| Db       | 9462   | TCTCGCCAGAGGGCTTCTCCTCAAGGTGTGCTCCAGACACCGACCTGCAAGGGGAT           | 9521 |
| QY       | 3073   | ValAlaGlyValAlaSerArgProSerAlaTrpArgSerGlyProValGlyAspProse        | 3093 |
| Db       | 9522   | GGCAGCTGGACAGGCAAGCCGCCCATCTGCTGAGAGGTCCGGCCCAATGGAGACCATC         | 9581 |
| QY       | 3093   | rThrLeuProGlySerHisArgSerProLysPro 3104                            |      |
| Db       | 9582   | AACACTGCCCCGGAGCCACCGCTACCCACCAAGCTT 9615                          |      |
| RESULT 9 |  |  |      |
| AS64376  |  |  |      |
| ID       | AS64376  | standard; DNA; 8010 BP.  |      |
| XX       | AS64376;                                       |  |      |
| XX       | 15-NOV-2002                                    | (first entry)  |      |
| DE       | Human cub and sushi domain containing gene #2. |  |      |

XX Human; NOVX; neurodegenerative disease; Alzheimer's disease; anxiety;  
 KM Parkinson's disease; Huntington's disease; neurological disorder;  
 KM schizophrenia; manic depression; mental retardation; angina pectoris;  
 KM cardiovascular disease; acute heart failure; myocardial infarction;  
 KM muscular disease; muscular disorder; retinal disease; photoreception;  
 KM death; keratinization disorder; inflammatory disease; ovarian cancer; melanoma;  
 KM immunological disorder; fungal infection; immune disease; diabetes;  
 KM bacterial infection; reproductive system disorder; metabolic disturbance;  
 KM viral infection; chronic disease; infectious disease;  
 KM anorexia; wasting disorder; chronic disease; infectious disease;  
 KM dyslipidaemia; cnd; sushi; myelin; von willebrand factor; kielin;  
 KM semaphorin; serine/threonine protein kinase; TGF-beta binding;  
 KM mas proto-oncogene; ribonuclease pancreatic precursor; aminotransferase;  
 KM tollid-like 2; cysteine sulfenic acid decarboxylase; gene; ds.  
 XX Homo sapiens.  
 XX WO20026791-A2.  
 XX 22-AUG-2002.  
 XX 10-DEC-2001; 2001WO-US048369.  
 XX 08-DEC-2000; 2000US-0254329P.  
 XX 14-DEC-2000; 2000US-0255648P.  
 XX 15-MAY-2001; 2001US-0291037P.  
 XX 08-JUN-2001; 2001US-0297173P.  
 XX 08-JUN-2001; 2001US-0309258P.  
 XX 29-AUG-2001; 2001US-0315639P.  
 XX 01-OCT-2001; 2001US-0326393P.  
 XX (CURA-) CRAGEN CORP.  
 XX Albrook JP, Anderson DW, Burgess CE, Boldog FL, Casman SJ;  
 PI Colman SD, Edinger SR, Ellerman K, Gerlach V, Gorman L, Grose WM;  
 PI Guo X, Hermann JL, Kikuda R, Lepley DM, Li L, Macdonnell JR;  
 PI Millet I, Pena CE, Peyman JA, Rastelli L, Rieger DK, Shinkets RA;  
 PI Smithson G, Szytek KA, Stone DJ, Tchernev VT, Vermet CAM, Voss EZ;  
 PI Zehnunen BD, Zhong H, Zhong M;  
 XX WPI; 2002-643486/69.  
 DR P-PSDB; ABG79169.  
 XX New NOVX polypeptides and polynucleotides useful for treating or  
 PT preventing e.g. neurodegenerative diseases, neurological disorders,  
 PT cardiovascular diseases, muscular diseases and disorders, or  
 PT immunological diseases.  
 XX Claim 9; Page 14-16; 299pp; English.  
 XX The present invention relates to new NOVX polypeptides. The polypeptides,  
 CC polynucleotides and antibodies are useful in the manufacture of a  
 CC medicament for treating or preventing neurodegenerative diseases (e.g.  
 CC Alzheimer's disease, Parkinson's disease, or Huntington's disease).  
 CC neurological disorders (e.g. anxiety, schizophrenia, manic depression or  
 CC mental retardation), cardiovascular disease (e.g. acute heart failure,  
 CC angina pectoris or myocardial infarction), muscular diseases and  
 CC disorders, retinal diseases (including those involving photoreception,  
 CC deafness and keratinization disorders), cancer (e.g. ovarian cancer or  
 CC melanoma), immunological disorders, inflammatory and immune diseases,  
 CC bacterial, fungal, protozoal and viral infections, and reproductive  
 CC system disorders. The proteins of the invention may be used to screen  
 CC drugs or compounds that modulate the NOVX protein activity or expression,  
 CC as well as to treat disorders characterised by insufficient or excessive  
 CC production of NOVX protein or protein forms that have decreased or  
 CC aberrant activity compared to NOVX wild type protein, such as diabetes,  
 CC obesity, metabolic disturbances associated with obesity, anorexia and  
 CC wasting disorders associated with chronic diseases and various cancers,  
 CC infectious diseases and various dyslipidemias. The nucleic acid  
 CC sequences of the invention may be used in chromosome mapping, identifying  
 CC an individual from minute biological samples (tissue typing), and in  
 CC forensic identification of a biological sample. The present nucleic acid

CC sequence encodes a NOVX protein of the invention  
 XX  
 SQ Sequence 8010 BP; 1720 A; 2412 C; 2140 G; 1738 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. NO: 0  
 Score: 14142.50  
 Percent Similarity: 97.71%  
 Best Local Similarity: 97.68%  
 Query Match: 83.26%  
 DB: 6  
 Gaps: 3  
 US-10-016-248-2 (1-3104) x ABS64376 (1-8010)  
 QY 1 MetAlaGlyAlaProProProAlaLeuLeuLeuProCysSerLeuIleSerAspCysCys 20  
 DB 1 ATGGCGGGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 60  
 QY 21 AlaSerArgGlnAlaGHisSerValGlyValGlyProSerGluLeuValIleGlnIle 40  
 DB 61 GCTAGCAATCAGACGACCTCCGTGGCGTAGGACCTCCGAGCTAGTCAGAGCAAAATT 120  
 QY 41 GluLeuIleSerArgGlyValIleValLeuMetProSerIleAspAsnSerGlnIleSer 60  
 DB 121 GAGTTGAAGTCTCGAGGTGTGAAGCTGATGCCAGCAAGCAACAGCAGCAAGCAAGCTCT 180  
 QY 61 ValLeuThrGlnValGlyValSerGlnIleHisAsnMetCysProAspProGlyIlePro 80  
 DB 181 GTGTAACTCAGGTGTGTGTCTCCCAAGCAATATATGTCTCAGACCTCGCATACCC 240  
 QY 81 GluArgGlyValArgLeuGlySerAspPheArgLeuGlySerSerValGlnPheThrCys 100  
 DB 241 CAAGGCGGCAAGAGCTAGGCTCGAATTCAGTTAGATCAGCGCCAGCTTCACTGTC 300  
 QY 101 ArgGlnIleIleIleThrLeuGlnIleSerIleValIleThrCysMetIleValSerAspMet 120  
 DB 301 AACGAGGCGCTATGACCTTCAGAGGCTCCAGGAGATACCTGTATGAAAGTGAAGCAATG 360  
 QY 121 PheAlaIleIleProSerAspPheIleArgProValCysArgAlaArgMetCysAspAlaHisLeu 140  
 DB 361 TTGGAGGCTGAGGACGACACAGGCGAGTCCGACCCGCAATGTGTGATGCCACCTT 420  
 QY 141 ArgGlyProSerGlyIleIleThrSerProAspPheProIleGlnIleIleIleIleIle 160  
 DB 421 CGAGGCGCCCTGGGCAATCATCACTCCCAATTTCCCATTTGATGACCAACAAATGCA 480  
 QY 161 HisCysValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 180  
 DB 481 CACTGTGTGTGATCATCAGACGACTCAACCCCTCCAAAGGTGATCAAGCTCGCTTGA 540  
 QY 181 GluPheAspLeuGluArgGlyValIleThrLeuThrValGlyValGlnIleIleIleIle 200  
 DB 541 GAGTTTATTTGAGAGGAGGCTTATGACACCTCAAGCTGCTGATGTGTGAGATGG 600  
 QY 201 ArgGlnIleThrValLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIle 220  
 DB 601 GACCAAGAGCAAGTCTCTATGATGTCAAAATAGCTGAGATGACAGCCCTCACCCCA 660  
 QY 221 GluSerArgIleProGluSerMetSerGlyAspIleIleIleIleIleIleIleIle 240  
 DB 661 GCTCTGCAATCCAGAGGAGCAAGTGTGGAGATTTGGAGGCGAGAAATGAGCTGACTT 720  
 QY 241 GluIleCysArgAspIleSerSerSerSerAspAlaArgSerGlySerValArgIleSerPro 260  
 DB 721 GAGATCTGTGTGATCACTTGAAGTTAGATGCAAGTCAAGTTCAGTGAAGAGTCTTCCA 780  
 QY 261 IyethSerAsnAlaValGluLeuValAlaProGlyThrGluIleGlnIleIleIleIle 280  
 DB 781 AATCTTCAATGCTGTGAACTTGTCTCTGCGGACAGATGAGAGGAGGAGTTC 840  
 QY 281 GlyAspProGlyIleProAlaIleIleIleIleIleIleIleIleIleIleIleIleIle 300  
 DB 841 GGTGACCTGTGCAATCTGATATGAGCGAGAGGAGGCTCCCGATTTCACCAAGGTGAC 900

```
QY 301 ThrLeuYsPheGluCySGluProAlaPheGluLeuValGlyGlnYsAlaIleThrCys 320
D 901 ACACCTCAAGTTGAGTGGACCCGCTTTAGCTGGTGGACAGAAAGGCAATACATCC 960
QY 321 GlnYsAanAngIntPSeAlaYsYsPheGluCysValPheSerCysPheAasn 340
D 961 CAATAAGATACCAATGGTGGCTTAAGAACGACGCTGGTCTCTGGTTCTTCAAC 1020
QY 341 PheThrSerProSerGlyValValLeuSerProAenTYrProGluAapTYrGlyAasnHis 360
D 1021 TTCACCAAGCCGCTGGGGTGTCTGCTCTCCAACTACCAAGAGGACTAAGGCAACAC 1080
QY 361 LeuHisCysValTYrPheLeuIleLeuAlaArgProGluSerArgIleHisLeuAlaPheAasn 380
D 1081 CTCACAGTGTCTGGCTCATCTGGCCAGGCTGAGAGCCGCAATCCACTGGCTTCAAC 1140
QY 381 AspIleAspValGluProGluPheAspPheLeuValIleYsAspGlyAlaThrAlaGlu 400
D 1141 GACATTGACGTGGAGCCTCACTTTGATTCTGGTCATCAAGATGGGGCCACCGCGAG 1200
QY 401 AlaProValLeuGlyYThrPheSerGlyAasnGlnLeuProSerSerIleThrSerSerGly 420
D 1201 GCGCCCGCTCGGCGACCTTCTCAGAAACCAAGCTTCCCTCCATCACAAAGAGTGGC 1260
QY 421 HisValAlaArgLeuGluPheGluThrAspHisSerThrGlyYsArgGlyPheAasnIle 440
D 1261 CACCTGGCCCTCTCGAGTTCAGACCTGACCACTCCAGGGAGAGGGGCTTCAACATC 1320
QY 441 ThrPheThrThrPheArgHisAasnGluCysProAspProGlyValProValAasnGlyYs 460
D 1321 ACTTTTACCACTTCCGACCAAGAGTGGCCGATCTGGCTTCCAGTAAATGGGAAA 1380
QY 461 ArgPheGlyAspSerLeuGlnLeuGlySerSerIleSerPheLeuCysAspGluGlyPhe 480
D 1381 CGGTTTGGGGACACCTTCCAGCTGGGCGCTCCACTCTTCTCTGTGATGAAGGGCTTC 1440
QY 481 LeuGlyThrGlnGlySerGluThrIleThrCysValLeuYsGluGlySerValValTyr 500
D 1441 CTTCGGACTCAGGGCTCAGAACCACTACCTGGCTTGAAGGAGGGAGGGTGTCTGG 1500
QY 501 AasnSerAlaValLeuArgCysGluAlaProCysGlyGlyHisLeuThrSerProSerGly 520
D 1501 AACAGCCCTGTGCTGGGCTGGAGCTCCCTGTGTGTGTACCTGACTGCCCAAGGGC 1560
QY 521 ThrIleLeuSerProGlyYThrProGlyPheTYrYsAspAlaLeuSerCysAlaTyrVal 540
D 1561 ACCATCTCTCTCCGGGCTGGCTGGCTTCAAGATGCTTGAAGCTGTGGGCTG 1620
QY 541 IleGluAlaGlnProGlyYThrProIleYsAlleThrPheAspArgPheYsThrGluVal 560
D 1621 ATTAGAGCCCAAGCAGGCTACCCATCAAAATCACTTCGACAAATTCAAAACGAGAGTC 1680
QY 561 AasnTYrAspThrLeuGluValAlaArgAspGlyYArgThrTYrSerAlaProLeuIleGlyVal 580
D 1681 AACATATACACCTCGAAGTACGCGAGTGGGGGCACTTACTCAGGGCCCTTGATGGGGTT 1740
QY 581 TYrHisGlyThrGlnValProGlnPheLeuIleSerThrSerAenTYrLeuTYrLeuLeu 600
D 1741 TACACCGGAGCCAGGTTCCCACTTCTCATCAGACACCACTACCTTCACTCTC 1800
QY 601 PheSerThrAspYsSerHisSerAspIleGlyPheGlnLeuAagTYrGluThrIleThr 620
D 1801 TTCTCTACCAAGCAAGATCATTCGACATCGGCTTCAGCTCCGCTATGAGACTATATACA 1860
QY 621 LeuGlnSerAspHisCysLeuAspProGlyYIleProValAasnGlyGlnArgHisGlyAasn 640
D 1861 CTGAGTCAAGACCACTGTCTGATCCAGGAATCCCAAGTAAATGACACGCTCATGGGAAT 1920
QY 641 AspPheTYrValGlyAlaLeuValThrPheSerCysAspSerGlyTYrThrLeuSerAsp 660
D 1921 GACTTCTACGTGGCGGCTGGTGAACCTTCAGTGTGACTGGGGCTACACATTAAGTAC 1980
QY 661 GlyValProLeuGluCysGluProAsnPheGluIntPSeArgAlaLeuProSerCysGlu 680
D 1981 GGGAGGCTCTGAGTGTGAGCCCACTTCCAGTGGAGCCGGGCTTGGCCAGTTGTGA 2040
QY 681 AlaLeuCysGlyGlyPheIleGlnGlySerSerGlyYThrIleLeuSerProGlyPhePro 700
D 2041 GCTCTGTGTGGCTTCATTCAAGGCTCCAGTGGAGACATCTTGTGGCAGGGTTCCT 2100
QY 701 AspPheTYrProAenAanLeuAasnCysThrTyrIleIleGluThrSerHisGlyYsGly 720
D 2101 GACTTCTACCCCAACCACTTGAACCTGCACTGTGATTCGAACATCTCATGGCAAGGT 2160
QY 721 ValPhePheThrPheHisThrPheHisLeuGlyYsSerGlyHisAspTYrLeuLeuIleThr 740
D 2161 GTGTTCTTCACTTCCACACTTCCACTGGAAGTGGCAATGACTCTCTCATACT 2220
QY 741 GluAasnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760
D 2221 GAGAACGGCAGCTTCAACCCAGCCCTGAGGCAAGTAACTGGATCTCGCTGGCAGCTCC 2280
QY 761 IleSerAlaGlyLeuTYrGlyAasnPheThrAlaGlnValArgPheIleSerAspPheSer 780
D 2281 ATCAGCGCTGGGCTCTATGGCACTTCACTGCCCCAGGTCCGCTTCATCTGTGATTTCTC 2340
QY 781 MetSerTYrGluGlyPheAasnIleThrPheSerGluTYrAspLeuGluProCysGluGlu 800
D 2341 ATGTCAATATAGAGATTCACATCACTTTCAGAGTACCACTTGGAGCCCTGTGAGAG 2400
QY 801 ProGluValProAlaTYrSerIleArgYsGlyLeuGlnPheGlyValGlyAspThrLeu 820
D 2401 CCGAGGCTCCAGCTACACATCCGGAAGGGCTTGGAGTTTGGCGTGGCGACACCTTG 2460
QY 821 ThrPheSerCysPhePheProGlyYThrArgLeuGluGlyYThrAlaArgIleThrCysLeuGly 840
D 2461 ACCTTCTCCGTCTCCCGGGTACCGTCTGAGGGCACCCCGCATCAAGTGTCTGGG 2500
QY 841 GlyArgArgArgLeuTPSeSerProLeuProArgCysValAlaGluCysGlyAasnSer 860
D 2521 GGCAGACGGGCTGTGGAGCTGCTCTGCTCCAGAGTGTGTGTGTGTGGGAAATTC 2580
QY 861 ValThrGlyThrGlnGlyYThrLeuLeuSerProAasnPheProValAenTYrAasnAasn 880
D 2581 GTCAAGAGCACTAGGTTACTTGTGTGTCTCCCACTTCTCTGTGAATCAATACAT 2640
QY 881 HisGluCysIleTYrSerIleGlnThrGlnProGlyYsGlyIleGlnLeuYsAlaArg 900
D 2641 CATGAATGATCTTACTTCATCCAGACCAAGGAGGAAATTCAGCTGAAGAGCCAGG 2700
QY 901 AlaPheGluLeuSerGluGlyAspValLeuYsValTYrAspGlyAasnAanSerAla 920
D 2701 GCATTCGAACTCTCCGAAGAGATGCTCAAGGTTTATATGGCAACAACATCCGCC 2760
QY 921 ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAanSerThrSer 940
D 2761 CGTTTGTGGAGATTTTATCCATTCATTCAGATATAGGGGGTGAATTTGAACAGACATCC 2820
QY 941 SerSerLeuTYrLeuAspPheIleThrAspAlaGluAanThrSerYsGlyPheGluLeu 960
D 2821 AGCAGTCTGTGGCTTGTATTCATCACTGAGTGAAGAAACACAGCAAGGGCTTGAACCTG 2880
QY 961 HisPheSerSerPheGluLeuIleYsCysGluAspProGlyYThrProYsPheGlyTYr 980
D 2881 CACTTTCACAGCTTGAATCATCAATGAGAGACCCAGAAACCCCAAGTTTGGCTAC 2940
QY 981 LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCysAspProGly 1000
D 2941 AAGTTTCATATAGAGTCAATTTTGCAGGAGAGCTCGTGTCTTCACTGCTGACCTCGA 3000
QY 1001 TYrSerLeuArgGlySerGluGluLeuLeuCysLeuSerGlyGluArgArgTYrTyrAsp 1020
D 3001 TACAGCTCTCGGGGTAGTGAAGAGCTGTGTGTGATGAGAGCGCCGAGCTTGGAG 3060
QY 1021 ArgProLeuProThrCysValAlaGluCysGlyYThrValArgGlyGluValSerGly 1040
```

```
Db 3061 CGGCTCTGCGCCACCTGTCCTCGCGAGTGTGAGGAGGACGTGAGAGAGAGGTGTGGGG 3120
Qy 1041 GlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIleTyrThr 1060
Db 3121 CAGGTCCTGTCAACCCGGGATACAGCTCCCTATGAACAACATCTCACTGCATCTGGACC 3180
Qy 1061 IleguaIagIuaIagIyCetherIlegIyLeuHisAspLeuValPheAspThrGluGlu 1080
Db 3181 ATCGAAGCAGAGGCGCGCTGCACCTTGGGCTACACTTCCTGGTGTGGACACAGAGAG 3240
Qy 1081 ValHisAspValLeuArgIleTrrAspGlyProValGluSerGlyValLeuLeuGlu 1100
Db 3241 GTTCAAGACGTCTCGCATCTGGAGTGGGCTGTGGAGAGCGGGGTTCTCTGAAGAG 3300
Qy 1101 LeuSerGlyProAlaLeuProIyAspLeuHisSerThrPheAsnSerValValLeuGln 1120
Db 3301 CTGATGGCGCCGCGCTCGCCAGGAGACCTGCATACACTTCACCTCGTGTCTGACAG 3360
Qy 1121 PheSerThrAspPhePheThrSerIySGlnGlyPheAlaIlegIlnPheSerValSerThr 1140
Db 3361 TTCACACCTGACTTCTTCACGACGAGAGGAGGCTTGGCATTCATTTTCAGTTCACA 3420
Qy 1141 AlaThrSerCysAsnAspProGlyIleProGlnAsnGlySerArgSerGlyAspSerTyr 1160
Db 3421 GCAACGTCTGCATGACCTCGGGATCCCGCAGATGGGAGTCCGAGTGGTGAAGATTGG 3480
Qy 1161 GluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaGlnGlnIySerAla 1180
Db 3481 GAGCGCGCGACTCCACAGCTTTCAGTGTACCTGGCTACCGGCTGCAGGAGAGTGA 3540
Qy 1181 GluIleSerCysValIySileGluAsnArgPhePheTrrGlnProSerProTrrCys 1200
Db 3541 GAGATCACTGTGGAAGATGAGAACAGGTTCTTGGCAGCCCAACCCCGCCACATGC 3600
Qy 1201 IleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr 1220
Db 3601 ATCCTCTCTGCGGGGAGACCTGACAGAACCATCTGAGATCACTCTCCACCAAAATTAC 3660
Qy 1221 ProGluProTyrProProGlyIyGluCysAspTrrPheValThrValSerProAsnTyr 1240
Db 3661 CCAGAACCTTACCCGAGGAGGAGAGTGTGACCTGGAAAGTACCGTCTCCACAGACTAC 3720
Qy 1241 ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle 1260
Db 3721 GTCACTCCCTGGATTTAACATCTTTAACCTGAGCTGGCTATGACTCTCCATATTC 3780
Qy 1261 TyrAspGlyArgAspSerLeuSerProLeuIlegIySerPheTyrGlySerGlnLeuPro 1280
Db 3781 TACGACGAGGAGGAGCTCTCAGCCCTCTCATAGAAAGCTTCTATGGCTCCACAGCTCCA 3840
Qy 1281 GlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal 1300
Db 3841 GGCGCGATTGAAAGCAGCAGCAACAGCTCTTCTCGCTCCGCGAGCATCATCTGTGG 3900
Qy 1301 SerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCysPheAspPro 1320
Db 3901 AGCAATGCTGGCTTCTCTCATTCATACAGAAACCCCGGAGTCAAGTTTATCTCT 3960
Qy 1321 GlySerIleIyAsnGlyThrArgValGlySerAspLeuIySerSerValThr 1340
Db 3961 GGTTCATACAAAGACGAGCAGCGGGTGGGCTCGACCTGAAGCTGGGCTCCCTCCAC 4020
Qy 1341 TyrTyrCysHisGlyValTyrGlyValGluGlyThrSerThrLeuSerCysIleLeuGly 1360
Db 4021 TACTACTGCGACGGGGGCTTACAGAAAGTTGAGGACCTTCAGACCTGAGCTGCTGGGG 4080
Qy 1361 ProAspGlyIyAspProValTrrAsnAspProArgProValCysThrAlaProCysGlyIy 1380
Db 4081 CCTGATGGAGAGCCGTGTGGAACAATCCCGGCAAGCTGTGACAGCCCTGTGGGGGA 4140
Qy 1381 GlnTyrValGlySerAspGlyValValLeuSerProAsnTyrProGlnAsnTyrThrSer 1400
Db 4141 CAGTATGTGGTTCGACGAGAGTGTCTTGTCCCAACTACCCCAAGACTACACAGAT 4200
Qy 1401 GlyGlnIleCysLeuTyrPheValThrValProIyAspTyrValValPheGlyGlnPhe 1420
Db 4201 GGACAGATCTGCTTGTATTTTGTACTGTCGCCAAGAGACTATGTGTGGCTGCTTC 4260
Qy 1421 AlaPhePheHisThrAlaLeuAsnAspValValGluValHisAspGlyHisSerGlnHis 1440
Db 4261 GCCTCTTTTCAACAGGCGCTCAACGACGTGGAGGTTCAAGACGCGCCACAGCCAGC 4320
Qy 1441 SerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGluSerLeuProLeuAlaThr 1460
Db 4321 TCGCGCTCTCAAGCTTCTCTCGGAGCTCCATACAGAGAAATCACTGCTTGGCCACC 4380
Qy 1461 SerAsnGlnValLeuIleIyAspPheSerAlaIySGlyLeuAlaProAlaArgIyPheHis 1480
Db 4381 TCCAAATCAAGTTCTCATTTAAGTTACGCCGCAAGGCTTCGCACCAAGAGGCTTCAC 4440
Qy 1481 PheValTyrGlnAlaValProArgThrSerAlaThrGlnCysSerSerValProGluPro 1500
Db 4441 TTTGTCTACAAAGCGTTCTCTGAACAGCGCCAGAGTGAAGCTGTGTCGGAAACC 4500
Qy 1501 ArgTyrGlyIyAspArgLeuGlySerAspPheSerValGlyValIleValArgPheGluCys 1520
Db 4501 CGCTATGGCAAGAGGCTGGCAGTACCTTCGGGTGGGGCCATCGCTTCGAATGC 4560
Qy 1521 AsnSerGlyTyrAlaLeuGlnGlySerProGluIlegIyCysLeuProValProGlyVala 1540
Db 4561 AACTCGGCTAGCCCTCAGAGGGGTCCGACAGATCGAGTCCCTCTGCTGCGGGCC 4620
Qy 1541 LeuAlaGlnTrrAsnValSerAlaProThrCysValValProCysGlyIyAsnLeuThr 1560
Db 4621 TTGGCCCAATGGAATGTCTCAGCGCCCAAGTGTGGCGCTGTGGAGGCAACTCACA 4680
Qy 1561 GluArgArgGlyThrIleLeuSerProGlyPheProGluProTyrIleAsnSerLeuAsn 1580
Db 4681 GAGCGCAGGGGACCATCTCTCCCTGGCTTCCACAGCCGTACCTCAACAGCTCAAC 4740
Qy 1581 CysValIleTrrPheIleValValProGluGlyValaGlyIlegIlnIlegIlnValSerPhe 1600
Db 4741 TGTGTGGAAGATCGTGTCCCGAAGGCGCTGCATCCAGATCCAAAGTTGTCAAGTTT 4800
Qy 1601 ValTrrGluGlnAsnTrrAspSerLeuGluValPheAspGlyValaAspAsnThrValThr 1620
Db 4801 GTGACAGAGCAACTGGAGCTCCTGGAAGTATTTAGGTGAGATTAACACTGTACC 4860
Qy 1621 MetLeuGlySerPheSerGlyThrThrValProAlaLeuAsnSerThrSerAsnGln 1640
Db 4861 ATGCTGGGAGATTCTCAGAAACACCGTGCCTGCTTGAACACACACTCCAACACAG 4920
Qy 1641 LeuTyrIleuHisPheTyrSerAspIleSerValSerAlaAlaGlyPheHisIleGluTyr 1660
Db 4921 CTCTACCTCATTTCTACTAGATACAGGATCTGCACCTGCTTCCACTTCCACTTGGAGTAC 4980
Qy 1661 LysThrValGlyLeuSerSerCysProGluProAlaValProSerAsnGlyValIySthr 1680
Db 4981 AAAACGGTGGGCTGAGCACTTGTCCGGAACCTGTGTGCCAGTAAAGGGGTGAAGACT 5040
Qy 1681 GlyGluArgTyrIleuValAsnAspValValSerPheGlnCysGluProGlyTyrAlaLeu 1700
Db 5041 GCGCAGGCTACTTGTGAATGATGTGTCTTTCAGAGTGAAGCGGGATATGCCCTTC 5100
Qy 1701 GlnGlyHisAlaHisIleSerCysMetProGlyThrValArgArgTrrAsnTrrProPro 1720
Db 5101 CAGGCGCACCGCCACATCTCTCAATGCCGGAACAGTGGGAGATGAAGAACTACCTCT 5160
Qy 1721 ProLeuCysIleAlaGlnCysGlyIyThrValGluGluMetGluGlyValIleLeuSer 1740
Db 5161 CCACTGTGATTCACAGTGTGGGGAACAGTGAAGATGAAGGGGTTGATCTCGAAGC 5220
Qy 1741 ProGlyPheProGlyIyAsnTyrProSerAsnMetAspCysSerTrrPheIleAlaLeuPro 1760
Db 5221 CCGGCTTCCAGGCAACTTACCCCAAGTACATGAGTGTCTCTGGAATAATAGCACTGCC 5280
```

|    |      |  |      |
|----|------|--|------|
| QY | 2121 | SerGIuYvSGInYrAaSGluPhaGluilePhekaSPGlyPProSeSGlyVgISeProleu       | 2140 |
| Db | 6361 | AGCGAGACCAATATGATGAAGTTTGAGATTTTATGAGTCCATAGACAGACAGTCTCTG         | 6420 |
| QY | 2141 | LeuYsaIaleuSeSGIYAsnTYrSerAlaProleuileValIThrSeSerSeSaSnSer        | 2160 |
| Db | 6421 | CTGAAGGCCCTCAGTGGGAATTACTCAGCTCCCTGATTTGTACACAGCTCAAGCACTCT        | 6480 |
| QY | 2161 | ValTYrleuAATGTTPSeSerAspHisAlaTYrAsnArglySGlyPheIysIlaArgTYr       | 2180 |
| Db | 6481 | GTTGATACCTCGCTTGATCTCATCTGATCAGCGCTACATGGAAGGGGCTTCAGATCCGCTAT     | 6540 |
| QY | 2181 | SerIlaProTYrCySeSerleuProArgAlaProleuHisGlyPheileuGlyGlnIThr       | 2200 |
| Db | 6541 | TCAGCCCTTATCTACAGCTGCGCCAGGGGCTCAGCTCCATGGCTTCATCTTAGGCCAGACC      | 6600 |
| QY | 2201 | SerThrgInPProGlyGlySerIleHisPheGlyCySaSnAlaGlyTYrArgIleuValGly     | 2220 |
| Db | 6601 | AGCACCCAGCCCGGGGGCTCCATCATCTTGCTGGTCAACGGCGGCTACCGCTGTGGGA         | 6660 |
| QY | 2221 | HisSerMetAlaIleCySerThrArgHisProGlnGlyTYrHisleuITPSeSerGluAlaIle   | 2240 |
| Db | 6661 | CACAGCATGGCCATCTGTACACCCGGCACCCCGAGGGTACACACTGTGGAGCGAAGCATC       | 6720 |
| QY | 2241 | ProleuCySeGlnAlaIleuSeSerCyGlyIleuProGluAlaProIlySaSnGlyMetValPhe  | 2260 |
| Db | 6721 | CCTCTGTGTCAAGGCTCTTCTGTGGGCTTCTTAGGCCCCCAAGATGGAATGATGTGTTT        | 6780 |
| QY | 2261 | GlyIySGIuTYrTrpValAGlyThrIlySaIaValTYrSeCySeSerGluGlyTYrHisIleu    | 2280 |
| Db | 6781 | GGCAAGAGATACAGATGGGAACCAAGGCCATGTACAGCTCAGTGAAGGCTTACCACTC         | 6840 |
| QY | 2281 | GlnIlaGlyAlaGluAlaThrAlaIleGlyCySaIeuAspThrGlyleuITPSeSaSnArgSn    | 2300 |
| Db | 6841 | CAGCAGGGCGGTGAGGCCACTGCAGAGTGTCTGACACAGGCTATGAGAGCAACGGCAT         | 6900 |
| QY | 2301 | ValIProProGInCySeVal-----ProValIThrCySePro                         | 2311 |
| Db | 6901 | GTCCACACACAGTGTGTCCGTGAGTCTCCGGCAATGAGAGCGGGGTGTGTGACTGTCTCT       | 6960 |
| QY | 2312 | AspValSerSerIleSerValGluHisGlyAArgTrpArgIleuIlePheGluThrGlnTYr     | 2331 |
| Db | 6961 | GATGTCAGTACATCAGCGTGGAGACATGGCCGATGGAGGCTTATCTTTGAGACACAGTAT       | 7020 |
| QY | 2332 | GlnPheGlnIaGlnIleuMetIleuIleCySaSPProGlyTYrTYrTYrThrGlyGlnArg      | 2351 |
| Db | 7021 | CAGTTCCAGGCCACGCTGATGCTCATCTGTGAGCCCTGGCTACTACTATACATGCGCCAAAG     | 7080 |
| QY | 2352 | ValIleArgCySeGlnAlaAsnGlyLYrITPSeSerIleuGlyAspSerThrProThrCySeArg  | 2371 |
| Db | 7081 | GTCATCTCGCTGTCAAGGCCAAATGGCCAAATGGAGGCTCCGGGAGACTCTACGCCCACTCGCGGA | 7140 |
| QY | 2372 | IleIleSerCySeGlyIleuIleuProIleProProAsnGlyHisArgIleGlyThrIleuSer   | 2391 |
| Db | 7141 | ATCATCTCTCTGTGAGAGCTCCGCAATTCGCCCAATGGCCACCGCATGGGAACACTGTCT       | 7200 |
| QY | 2392 | ValTYrGlyAlaThrAlaIlePheSeCySaSnSerGlyTYrThrIleuValAGlySerArg      | 2411 |
| Db | 7201 | GTCACGGGGGAACAGCATCTTCTCTCTGCAATTCGGGATACACACTGTGGGGCTCCAGG        | 7260 |
| QY | 2412 | ValIArgGlyCySeMetAlaAsnGlyIleuITPSeSerGlySerGluValArgCySeIleu----- | 2429 |
| Db | 7261 | GTCGTGATAGTCATAGGCCAAATGGGCTCTGGAGTGGCTCTGAAGTCCCGCTGCTTGCCACT     | 7320 |
| QY | 2429 | -----  | 2429 |
| Db | 7321 | CAGACCAAGCTCCACTCCATTTCTATTAAGCTCCTCTTCAGATGTACTCTTCCCATCC         | 7380 |
| QY | 2430 | -----AlaGlyHisCySeGlyThrProGluProIleValAsnGlyHisIleAsnGly          | 2446 |
| Db | 7381 | CTCACCAAAAGCTGACACTGTGGAGACTCTGAGGCCATTTGTCAACGGACACATCAATCAAGG    | 7440 |
| QY | 2447 | GluAsnTYrSerTYrArgIySerValValTYrGInCySaSnAlaGlyIlePheArgIleuIle    | 2466 |

|    |  |                  |  |
|----|--|------------------|--|
| PR | 07-JUN-2002  | 2002US-0386591P  |  |
| PR | 07-JUN-2002  | 2002US-0386592P  |  |
| PR | 07-JUN-2002  | 2002US-0386597P  |  |
| PR | 07-JUN-2002  | 2002US-0387262P  |  |
| PR | 08-JUN-2002  | 2002US-0296960P  |  |
| PR | 10-JUN-2002  | 2002US-0387400P  |  |
| PR | 10-JUN-2002  | 2002US-0387535P  |  |
| PR | 11-JUN-2002  | 2002US-0387610P  |  |
| PR | 11-JUN-2002  | 2002US-0387625P  |  |
| PR | 11-JUN-2002  | 2002US-0387634P  |  |
| PR | 11-JUN-2002  | 2002US-0387666P  |  |
| PR | 11-JUN-2002  | 2002US-0387696P  |  |
| PR | 11-JUN-2002  | 2002US-0387702P  |  |
| PR | 11-JUN-2002  | 2002US-0387836P  |  |
| PR | 11-JUN-2002  | 2002US-0387858P  |  |
| PR | 12-JUN-2002  | 2002US-0387934P  |  |
| PR | 12-JUN-2002  | 2002US-0387934P  |  |
| PR | 12-JUN-2002  | 2002US-0388020P  |  |
| PR | 12-JUN-2002  | 2002US-0388062P  |  |
| PR | 12-JUN-2002  | 2002US-0388096P  |  |
| PR | 13-JUN-2002  | 2002US-0389123P  |  |
| PR | 14-JUN-2002  | 2002US-0389118P  |  |
| PR | 14-JUN-2002  | 2002US-0389120P  |  |
| PR | 14-JUN-2002  | 2002US-0389144P  |  |
| PR | 14-JUN-2002  | 2002US-0389146P  |  |
| PR | 17-JUN-2002  | 2002US-0389729P  |  |
| PR | 17-JUN-2002  | 2002US-0389742P  |  |
| PR | 18-JUN-2002  | 2002US-0389848P  |  |
| PR | 19-JUN-2002  | 2002US-0390068P  |  |
| PR | 19-JUN-2002  | 2002US-0390209P  |  |
| PR | 21-JUN-2002  | 2002US-0390763P  |  |
| PR | 17-JUL-2002  | 2002US-0396706P  |  |
| PR | 06-AUG-2002  | 2002US-0401628P  |  |
| PR | 09-AUG-2002  | 2002US-0402156P  |  |
| PR | 09-AUG-2002  | 2002US-0402256P  |  |
| PR | 09-AUG-2002  | 2002US-0402389P  |  |
| PR | 12-AUG-2002  | 2002US-0402786P  |  |
| PR | 12-AUG-2002  | 2002US-0402816P  |  |
| PR | 12-AUG-2002  | 2002US-0402821P  |  |
| PR | 12-AUG-2002  | 2002US-0402832P  |  |
| PR | 13-AUG-2002  | 2002US-0403448P  |  |
| PR | 13-AUG-2002  | 2002US-0403459P  |  |
| PR | 13-AUG-2002  | 2002US-0403531P  |  |
| PR | 13-AUG-2002  | 2002US-0403532P  |  |
| PR | 13-AUG-2002  | 2002US-0403563P  |  |
| PR | 13-AUG-2002  | 2002US-0406317P  |  |
| PR | 15-AUG-2002  | 2002US-0406179P  |  |
| PR | 26-AUG-2002  | 2002US-0406182P  |  |
| PR | 26-AUG-2002  | 2002US-0406355P  |  |
| PR | 27-AUG-2002  | 2002US-0406240P  |  |
| PR | 12-SEP-2002  | 2002US-0410084P  |  |
| PR | 20-SEP-2002  | 2002US-0410288P  |  |
| PR | 23-SEP-2002  | 2002US-0412731P  |  |
| PR | 30-SEP-2002  | 2002US-0414801P  |  |
| PR | 30-SEP-2002  | 2002US-0414839P  |  |
| PR | 30-SEP-2002  | 2002US-0414840P  |  |
| PR | 30-SEP-2002  | 2002US-0414954P  |  |
| PR | 09-OCT-2002  | 2002US-0417466P  |  |
| PR | 09-OCT-2002  | 2002US-0417466P  |  |
| PR | 23-OCT-2002  | 2002US-0420639P  |  |
| PR | 28-OCT-2002  | 2002US-0421566P  |  |
| PR | 31-OCT-2002  | 2002US-0422690P  |  |
| PR | 01-NOV-2002  | 2002US-0423130P  |  |
| PR | 05-NOV-2002  | 2002US-00423798P |  |
| PR | 05-NOV-2002  | 2002US-0423798P  |  |
| PR | 12-NOV-2002  | 2002US-0425453P  |  |
| PR | (CURA-) CURAGEN CORP.  |                  |  |
| PR | Alsbrock JB, Alvarez E, Anderson DW, Boldog FL, Casman SU, Catterton E, Chapolval A, Crabtree-Bokor JR, Edinger SR, Eileman K, Gieney VZ, Gannoli EA, Gerlach VL, Gorman L, Gunther E, Guo X, Hermand H, Kohn J, W |                  |  |



QY ThrIleLeuSerProGlyTyrProGlyPheTyrIleuSerAlaLeuSerCysAlaTyrVal 540  
 DB ACCATCTCTCTCCGGGCTGAGCTTCAAGAGATGCTTGAAGTGGCTGGAGTG 1620  
 QY IleguaIagInProGlyTyrProIleuValIleThrPheAspArgPheValThrGluVal 560  
 DB ATTAGAGCCAGCCAGGCTTACCCCATCAAAATCAGCTTGCAGATTCAAACCGAGAGTC 1680  
 QY AsnTyrAspThrLeuGluValArgAspGlyArgThrTyrSerAlaProLeuIleGlyVal 580  
 DB AACTATGACACCTCGAAGTACGCGATGGCGGACTTACAGCGCCCTTCATCGGGGTT 1740  
 QY TyrIleGlyThrGluValProGluPheLeuIleSerThrSerAsnTyrLeuTyrLeuLeu 600  
 DB TACACCGGAGCCAGGTTCCCAAGTCTCTCATCAGCAGCAGCACTTACTTCTCTC 1800  
 QY PheSerThrAspIleSerHisSerAspIleGlyPheGluLeuArgTyrGluThrIleThr 620  
 DB TTCTCTACCGACAAGAGTCACTCGGATCCGCTTCCAGCTCCGCTATGAGACTATACA 1860  
 QY LeuGluSerAspHisCysLeuAspProGlyIleProValAsnGlyValAsnGlyValAsn 640  
 DB CTGACGTACAGCACTGCTTGATCCAGAAATCCAGTAATGACAGCGTCAATGGGAT 1920  
 QY AspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrThrLeuSerAsp 660  
 DB GACTTCTACGCGGCGCTGTCACCTTCACTGTCAGTCCGGCTACATTAAGTAC 1980  
 QY GlyGluProLeuGluCysGluProAsnPheGluTyrSerArgAlaLeuProSerCysGlu 680  
 DB GGGAGCCTCTGGAGTGGAGCCCACTTCCAGTGGAGCGGGCCCTGACAGTTGGAA 2040  
 QY AlaLeuCysGlyGlyPheIleGluIleSerSerGlyThrIleLeuSerProGlyPhePro 700  
 DB GCTCTCTGCTGGCTTCACTTCAAGGCTCCAGTGGAGCACTTGTGCGAGGATTCCT 2100  
 QY AspPheTyrProAsnAsnLeuAsnCysThrTyrIleIleGluThrSerHisGlyValGly 720  
 DB GACTTCTACCCCAACAATTGAACCTGCACTGGAATTCGAAACATCTCATGCGAAGGT 2160  
 QY ValPhePheThrPheHisThrPheHisAlaLeuGluSerGlyHisAspTyrLeuLeuIleThr 740  
 DB GTGTTCTTCACTTCCACACCTTCCACCTGAAAGTGGCATACCTTCCATCACT 2220  
 QY GluAsnGlySerPheThrGluProLeuArgGluLeuThrGlySerArgLeuProAlaPro 760  
 DB GAGAACCGGAGCTTCAACCCAGCCCTGAGGCACTAAGTGAATCTGGCTGCCAGCTCCC 2280  
 QY IleSerIleGlyLeuTyrGlyAsnPheThrAlaGluValArgPheIleSerAspPheSer 780  
 DB ATCAGCGCTGGGCTCTATGCGAACTTCACTGAGCCAGATCGGCTTCACTCTGATTTCTCC 2340  
 QY MetSerTyrGluGlyPheAsnIleThrPheSerGluTyrAspLeuGluProCysGlyGlu 800  
 DB ATGTCAATATGAAGATTCAACATCACTTCTCAAGATGAGACTTGGAGCCCTGTGAGG 2400  
 QY ProGluValProAlaTyrSerIleArgGlyLeuGluPheGlyValGlyAspThrLeu 820  
 DB CCCGAGGTCCAGCTTCAAGCATCCCGAAGGCGCTTCACTTGGCTGGCGGCACTTGG 2460  
 QY ThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIleThrCysLeuGly 840  
 DB ACCTTCTCTCTCTCCCGGAGTACGCTGAGAGGAGCAGCCGCGATCACTGCGCTGGGG 2520  
 QY GlyArgArgArgLeuTyrSerSerProLeuProArgCysValAlaGluCysGlyValAsnSer 860  
 DB GGCAGACGGGCGCTGAGAGCTGCTGCCAAGTGTGCTGAGTGGAGTGGAAATTC 2580  
 QY ValThrGlyThrGluGlyThrLeuLeuSerProAsnPheProValAsnTyrAsnAsnAsn 880  
 DB GTCAAGGAGCTCAAGGTACTTGTGCTGCCCACTTCTGTGAACATAATACAT 2640

QY HisGluCysIleTyrSerIleGluThrGluInProGlyValGlyIleGluLeuValArg 900  
 DB CATGAATGATCATCACTCCATCCAGACCCAGGAGGAGGATTCAGCTGAAGCCAG 2700  
 QY AlaPheGluLeuSerGlyValAspValLeuValValTyrAsnGlyAsnAsnAsnSerAla 920  
 DB GCATTCGAATCTCCGAGAGATGTCTCAAGTTTATGATGGCGCAACAACATCCGCG 2760  
 QY ArgLeuLeuGlyValPheSerHisSerGluMetGlyValThrLeuAsnSerThrSer 940  
 DB CGTTTCTGGAGATTTTTATGCCATTCGATGAGATGGGGGTGACTTGAACGACATCC 2820  
 QY SerSerLeuTyrLeuAspPheIleThrAspAlaGluAsnThrSerIleGlyPheGluLeu 960  
 DB AGAGTGTGTGCTGATTTCACTCATGATGCTGAACACCCAGCAGAGGCTTGAACAG 2880  
 QY HisPheSerSerPheGluLeuLeuIleCysGlyAspProGlyThrProValPheGlyTyr 980  
 DB CACTTTCAGCTTGAATCACTCAATGTGAGAACCCAGAACCCCAAGTTGGCTAC 2940  
 QY LysValHisAspGluGlyHisPheAlaGlySerSerValIlePheSerCysAspProGly 1000  
 DB AAGTTTCATGATGAAGTCACTTTTGCAGGAGCTCCGCTGCTTCACTGAGTGCCTG 3000  
 QY TyrSerLeuArgGlySerGluGluLeuLeuCysLeuSerGlyValArgThrTyrAsp 1020  
 DB TACAGCTGGGGAGTGAAGAGCTGTGTGTGATGAGAGAGAGCGCGGAGCTGGAG 3060  
 QY ArgProLeuProThrCysValAlaGluCysGlyGlyTyrValArgGlyValValSerGly 1040  
 DB CGGCTCTGCCCACTGTGTGCGCGAGTGGAGGACAGTGAAGAGAGAGTGTGGGG 3120  
 QY GlnValIleLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIleTyrThr 1060  
 DB CAGGTGCTGTACCCGAGTATCCAGCTCCATGAAACAATCTCACTGATCGAGC 3180  
 QY IleguaIagIuaIagIleCysThrIleGlyLeuHisAspLeuValPheAspThrGluGlu 1080  
 DB ATCGAAGAGAGAGCCGCTGACCACTTGGCTTACATCTCTGATGTTTGACACAGAGAG 3240  
 QY ValHisAspValLeuArgIleTyrAspGlyProValGluSerGlyValLeuLeuValGlu 1100  
 DB GTTCAAGAGCTGGCGCATCTGGAGTGGGCTGTGGAGAGCGGGTTCTGTCAAGAG 3300  
 QY LeuSerGlyProAlaLeuProValAspLeuHisSerThrPheAsnSerValIleGluGln 1120  
 DB CTGAGTGGCCGGCCCTGGCCCAAGACCTGATGCACTTCAACTGCGTGTCTGCGAG 3360  
 QY PheSerThrAspPhePheThrSerIleGlyPheAlaIleGluPheSerValSerThr 1140  
 DB TTCAGCACTGACTTCTTCAACAGCAAGAGGCTTTCCTCAATTTTCAAGTGTCCACA 3420  
 QY AlaThrSerCysAsnAspProGlyIleProGluAsnGlySerArgSerGlyAspSerTyr 1160  
 DB GCAACGTCTCGCATGACCTGTGGATCCCGAGATGGAGTGGAGTGTGACAGATTGG 3480  
 QY GluAlaGlyAspSerThrValPheGluCysAspProGlyTyrValAlaLeuGluGlySerAla 1180  
 DB GAACCCGCGCACTCCACAGTGTTCATGATGACCTTGGCTACCGGCTGCGAGGAGTGA 3540  
 QY GluIleSerCysValIleIleguaIagIuaIagIuaIagIuaIagIuaIagIuaIagIua 1200  
 DB GAGATCAGCTGTGTGAAGATCGAAGACAGGTTCTTGTGGAGCCGAGCCGCAACATGC 3600  
 QY IleAlaProCysGlyValAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr 1220  
 DB ATGCTCTCCCTGGGGAGAGCTGACAGACCACTCGAGTCACTCTCCACCAAAATTAC 3660  
 QY ProGluProTyrProProGlyValGluCysAspTyrValValThrValSerProAspTyr 1240  
 DB CAGAACCTCAACCCGCAAGGCAAGAGTGTGATGAAAGTGCAGCTTCAACAGACTAC 3720  
 QY ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle 1260

Db 3721 GTGATCGCCCTGGTATTAACTTTAACTGAGGAGCCTGGCTAATGACTTCTCCATATC 3780  
Qy 1261 TyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerGlnLeuPro 1280  
Db 3781 TACAGCGAGCGGAGCTCTCTCAGCCCTCTCATAGAACCTTCTATGGCTCCACAGTCCCA 3840  
Qy 1281 GlyArgIleGlySerSerAspSerPheLeuAlaPheArgSerAspAlaSerVal 1300  
Db 3841 GGGCGCATTTGAAAGACAGACAGACCTTCTCTCGCTTCCGAGGAGTACATCTGTG 3900  
Qy 1301 SerAsnAlaGlyPheValIleAspTyrThrGlnAsnProArgGlySerCysPheAspPro 1320  
Db 3901 AGCAATCGTGGCTTGTCTATTGACTATACAGAAAACCGCGGAGTCAATGTTTGTACT 3960  
Qy 1321 GlySerIleGlyAsnGlyThrArgValIleGlySerAspLeuGlySerSerValThr 1340  
Db 3961 GGTTCCATCAAGAACGAGCACACGGGTGGGTCCGACTGAAAGCTGGGCTCCCTCCATCAC 4020  
Qy 1341 TyrTyrCysHisGlyGlyTyrGluValIleGlyThrSerThrLeuSerCysIleLeuGly 1360  
Db 4021 TACTATCTCCACGGGGGCTTACGAGTTAGGGACCTGACCTGAGCTGACATCTGGGG 4080  
Qy 1361 ProAspGlyArgProValITrpAsnAsnProArgProValCysThrAlaProCysGlyGly 1380  
Db 4081 CCTGATGGGAAGCCCGTGTGGAAACATCCCGGCGCAGTCTGACAGCCCTGTGGGGGA 4140  
Qy 1381 GlnTyrValIleGlySerAspGlyValValLeuSerProAsnTyrProGlnAsnTyrThrSer 1400  
Db 4141 CAGTATGGGTGGTCCGAGCGAGGTGCTTGTCTCCCACTACCCCAACATACACCCAGT 4200  
Qy 1401 GlyGlnIleCysLeuTyrPheValIThrValProIleAspTyrValValPheGlyGlnPhe 1420  
Db 4201 GGACAGATCTGCTGTATTGTATTACTGTGCCAAGACATAGTGTGTTGGCCAGTTC 4260  
Qy 1421 AlaPhePheIleThrAlaLeuAsnAspValValIleValHisAspGlyHisSerGlnHis 1440  
Db 4261 GCCTTCTTTCAACGGGCTTCAACGACGTGGTGGAGGTTTCAAGACGCGCACAGCCAGCAG 4320  
Qy 1441 SerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGlySerLeuProLeuAlaThr 1460  
Db 4321 TCGGGGCTCTTCAGCTCCCTCGGGCTCCCATACAGAGATCACTCCCTTGGCCACC 4380  
Qy 1461 SerAsnGlyValLeuIleValPheSerAlaValGlyLeuAlaProAlaArgGlyPheHis 1480  
Db 4381 TCCATCAAGATTCTCATTAAGTTCAGCGCCAAAGGCTTCGCACCAAGCCAGGCTTCCAC 4440  
Qy 1481 PheValTyrGlnAlaValProArgThrSerAlaThrGlnCysSerSerValProGlnPro 1500  
Db 4441 TTGTGTCACCAAGCGGTTCTCTCGAACCAAGCCGACGCTGCTGTCCGGAAACC 4500  
Qy 1501 ArgTyrGlyArgArgLeuGlySerAspPheSerValGlyAlaIleValArgPheGlnCys 1520  
Db 4501 CGCTATGGCAAGAGGCTGGGAGTGACTTCTCGGGGGGCGCATGCTCGCTCGAATGC 4560  
Qy 1521 AsnSerGlyTyrAlaLeuGlnGlySerProGlnIleGlyCysLeuProValProGlyVal 1540  
Db 4561 AACTCGGCTATGCGCTGCAAGGGGTCCGACAGATCGAGTCCCTCCGTGCTCGGGGCC 4620  
Qy 1541 LeuAlaGlnITrPAsnValSerAlaProThrCysValValProCysGlyGlyAsnLeuThr 1560  
Db 4621 TTGGCCCAATGGAATGCTTACGCGCCACGCTGTGTGGTGGTGGAGGACACTCACA 4680  
Qy 1561 GlnArgArgGlyThrIleLeuSerProGlyPheProGlnProTyrLeuAsnSerLeuAsn 1580  
Db 4681 GAGCGCAGAGGGGACCATCTGTCTCCCTGCTCCAGAGCCGTAACCAAGGCTCAAC 4740  
Qy 1581 CysValITrPLeuValIleValProGlnGlyValGlyIleGlnIleGlnValIleSerPhe 1600  
Db 4741 TGTGTGTGAAGATGTGTGTGTCCGAAAGGCTGACATTCAGATCAAGTTGTCAAGTTT 4800  
Qy 1601 ValIThrGlnGlnAsnITrPAspSerLeuGlnValPheAspGlyAlaAspAsnThrValThr 1620

Db 4801 GTGACAGACGAACTGGGACTCGCTGAAGTATTGATGTGCAGATTAACACTGTAAAC 4860  
Qy 1621 MetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerAsnGln 1640  
Db 4861 ATGCTGGAGATTCTCAGAAACACCGTGTCCCTTCTGAAACACACACTTCAACACAG 4920  
Qy 1641 LeuTyrLeuHisPheTyrSerAspIleSerValSerAlaIleGlyPheHisIleLeuGlyThr 1660  
Db 4921 CTCACTTCACTTTCTACTCAGATATCAGGCTATCTGACACTGCTTCCACTTGGATAC 4980  
Qy 1661 LysThrValIleGlyLeuSerSerCysProGlnProAlaValProSerAsnGlyValIleThr 1680  
Db 4981 AAAACGGTGGGCTTGAACAGATTGTCGGAACCTGCTGTCCCAAGTAAACGGGGTGAAC 5040  
Qy 1681 GlnGlnArgTyrIleValIleAsnAspValIleSerPheGlnCysGlnProGlyTyrAlaLeu 1700  
Db 5041 GGCAGAGGCTCACTTGGTGAATGAATGTGTCTTCCAGTGTGACGGGAGTATGCTCCTC 5100  
Qy 1701 GlnGlyHisAlaHisIleSerCysMetProGlyTyrThrValITrPAsnTyrProPro 1720  
Db 5101 CAGGACCAACGCCCACTCTCTGCTGATGCCGGAACAGTGGGCGATGGAATCACTCTCT 5160  
Qy 1721 ProLeuCysIleAlaGlnCysGlyGlyThrValIleGlnLeuGlyValIleLeuSer 1740  
Db 5161 CCACTTGTATTTGACAGTGTGGGGGAAACGTGGAGAGATGGAGGGGTGATCTGAGC 5220  
Qy 1741 ProGlyPheProGlyValAsnTyrProSerAsnMetAspCysSerITrPLeuAlaLeuPro 1760  
Db 5221 CCCGCTTCCACAGGCAACTACCCAGTAAATGACATGCTCCGGAATAAAGCACTCC 5280  
Qy 1761 ValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrGlnProAsnHisAspTyr 1780  
Db 5281 GTGGGCTTTGGAGTCAATCACTCACTTCTGAACTTCTTCCACCGAGCCCAACAGCACTAC 5340  
Qy 1781 IleGlnIleArgAsnGlyProTyrGlnThrSerArgMetGlyArgPheSerGlySer 1800  
Db 5341 ATAGAAATCCGGAATGAGCCCTATGAGACACAGCCGATGAGGAGATTCAGTGAAGC 5400  
Qy 1801 GlnLeuProSerSerLeuLeuSerThrSerHisGlnThrValITrPheHisSerAsp 1820  
Db 5401 GAGCTTCAAGCTCCCTCTCTCCACGCTCCACAGACCAACCGTATTTTCCACAGGAC 5460  
Qy 1821 HisSerGlnAsnArgProGlyPheIleLeuGlnITyrGlnAlaTyrGlnLeuGlnGlyCys 1840  
Db 5461 CACTTCCAGATTCGCGCAGATTCAGCTGAGATATAGGCTTATGAATTCCTCAAGATGC 5520  
Qy 1841 ProAspProGlnProPheAlaAsnGlyIleValArgGlyValIleGlyTyrAsnValIleGln 1860  
Db 5521 CCAGACCCAGAGCCCTTTGGCAATGCGATTGTGAGGGAGCTGGCTCAACGTGGGACAA 5580  
Qy 1861 SerValIThrPheGlnCysLeuProGlyTyrGlnLeuThrGlyHisIleProValLeuThrCys 1880  
Db 5581 TCAGTGACTTTCGAGTCTCCCGGGGTATCAATTGACTGGCCACCTGTCTTCCACCTGT 5640  
Qy 1881 GlnHisGlyIThrAsnArgAsnITrPAsnHisProLeuProIleCysGlyValProCysGly 1900  
Db 5641 CAACATGACACCAACCGAACTGGAGCACCCCTGCGCAAGTGTGAATGCCCTGTGGC 5700  
Qy 1901 GlyAsnIleThrSerSerAsnGlyThrValITrSerProGlyPheProSerProTyrSer 1920  
Db 5701 GGGAACTATCACTTCTTCCACGACGCTGTACTCCCGGGGTTCCCTTACCCGTAATCC 5760  
Qy 1921 SerSerGlnAspCysValITrPLeuIleThrValProIleGlyHisGlyValIleGlnLeuAsn 1940  
Db 5761 AGCTCCACAGACGTGTCTGCTGATACCGGTGCCATTTGCCATGGCGTCCGCTCAAC 5820  
Qy 1941 LeuSerLeuGlnGlnITrGlnProSerGlyAspPheIleThrIleITrPAspGlyProGln 1960  
Db 5821 CTGAGCTGCTGACAGACAGCCCTCTGGAGATTTCAATCCATCTGGAGTGGGCCACAG 5880  
Qy 1961 GlnIThrAlaProArgLeuGlyValPheThrArgSerMetAlaValIleGlnSer 1980  
Db 5881 CAAACAGACACAGGCTCGGGCTTTCACCCGGAGCATGGCCAAAGAAAACAGTGCAGAGT 5940

|                                      |      |   |      |
|--------------------------------------|------|---|------|
| Qy                                   | 1981 | SerSerAsnGlnValLeuLeuPheHisArgAspAlaAlaThrGlyGlyIlePheHis       | 2000 |
| Db                                   | 5941 | TCATCCAAACAGAGTCTCTGCTCAAGTTCCACCGTGAATGACGACAGGAGGAGTCTTGCC    | 6000 |
| Qy                                   | 2001 | IleAlaPheSerAlaTyrProLeuThrIysCysProProProThrIleLeuProAsnAla    | 2020 |
| Db                                   | 6001 | ATAGCTTTCTCCGCTTATCCATCCCAATGCTCCTCCACCATCTCTCCCAAGCC           | 6060 |
| Qy                                   | 2021 | GluValIValThrGluAsnGluLupPheAsnIleGlyAspIleValArgTyrArgCysLeu   | 2040 |
| Db                                   | 6061 | GAAAGTCGACAGAAATGAAGAAATTAATATAGGTGACATCGATGACGATGACGCTC        | 6120 |
| Qy                                   | 2041 | ProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuGlyThrTyrLeuGln    | 2060 |
| Db                                   | 6121 | CTGGCTTACCTTAGTGGGGAATGAATTCGACCTGCAACCTTGAACTTACCTGAG          | 6180 |
| Qy                                   | 2061 | PheGluGlyProProProIleCysGluValHisCysProThrAsnGluLeuLeuThrAsp    | 2080 |
| Db                                   | 6181 | TTTGAAGAACCAACCCCGATATGTGAAGTGCATGTCACAAATGAGCTTCTGACAGAC       | 6240 |
| Qy                                   | 2081 | SerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCys    | 2100 |
| Db                                   | 6241 | TCGACAGGCGTGAATCTTACGACCAAGCTACCTCGAAGCTATCCCGAGTTCCAGACTGC     | 6300 |
| Qy                                   | 2101 | SerThrLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeu    | 2120 |
| Db                                   | 6301 | TCTTGCGCTGAGAGTGAGCCGACCTAATACATCTCCCTCAGAGTGAATCTTCTC          | 6360 |
| Qy                                   | 2121 | SerGluYsgGlnTyrAspGluPheGluIlePheAsnGlyProSerGlyGlnSerProLeu    | 2140 |
| Db                                   | 6361 | AGCGAAGAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT        | 6420 |
| Qy                                   | 2141 | LeuValAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerSerAsnSer    | 2160 |
| Db                                   | 6421 | CTGAAGGCTTCAGTGGGAATTAATCTCAGCTCCCTGATTTGACCACTCAAGCACTCT       | 6480 |
| Qy                                   | 2161 | ValTyrLeuArgTyrPheSerSerAspHisAlaTyrAsnArgLysGlyPheLysIleArgTyr | 2180 |
| Db                                   | 6481 | GTGTAACCTGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT       | 6540 |
| Qy                                   | 2181 | SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr    | 2200 |
| Db                                   | 6541 | TCAGCCCTTACGACACCTGCGACAGGCTCCACTCATGAGCTTCACTTCAAGCCAGACC      | 6600 |
| Qy                                   | 2201 | SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly    | 2220 |
| Db                                   | 6601 | AGCACCCAGGCGCGGCTCATCTCACTTGGCTGCAACGCGGCTACCGCTGAGGGA          | 6660 |
| Qy                                   | 2221 | HisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuThrPheSerGlnAlaIle | 2240 |
| Db                                   | 6661 | CACAGCATGGCCATCTTACCCGACACCCGAGGCTACCACTGAGGAGCGAAGCCATC        | 6720 |
| Qy                                   | 2241 | ProLeuCysGlnAlaLeuSerCysGlyLeuProGlnAlaProLysAsnGlyMetValPhe    | 2260 |
| Db                                   | 6721 | CTCTCTGTCAGACTCTTCTCTGAGGCTTCTGAGGCTTCTGAGGCTTCTGAGGCTT         | 6780 |
| Qy                                   | 2261 | GlyLysGluTyrThrValGlyThrValAlaValTyrSerCysSerGlyGlyTyrHisLeu    | 2280 |
| Db                                   | 6781 | GAGGAGGAGTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG         | 6840 |
| Qy                                   | 2281 | GlnAlaGlyValAlaGlnAlaThrAlaGluCysLeuAspThrGlyLeuThrSerAsnArgAsn | 2300 |
| Db                                   | 6841 | CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG         | 6900 |
| Qy                                   | 2301 | ValProProGlnCysVal-----ProValThrCysPro                          | 2311 |
| Db                                   | 6901 | GTCCCAACCAAGTGTCTCGTGAAGTCTCGGCAATGAGAGGCGGAGTGTGATCTGTCT       | 6960 |
| Qy                                   | 2312 | AspValSerSerIleSerValGluHisGlyArgTyrPheLeuIlePheGluThrGlnTyr    | 2331 |
| Db                                   | 6961 | GATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT        | 7020 |
| Qy                                   | 2332 | GlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyrTyrTyrThrGlyGlnArg    | 2351 |
| Db                                   | 7021 | CAGTTCAGAGCCCAAGCTGATGCTCATCTGTGACCTGGCTACTACTATCTGGCAAAAG      | 7080 |
| Qy                                   | 2352 | ValIleArgCysGlnAlaAsnGlyLysTrpSerLeuGlyAspSerThrProThrCysArg    | 2371 |
| Db                                   | 7081 | GTGATCCGCTGTAGGCGCAATGAGCAATGAGCTCTGGAGACTTACGCGCCACCTGCCA      | 7140 |
| Qy                                   | 2372 | IleIleSerCysGlyGluLeuProIleProProAsnGlyHisArgIleGlyThrLeuSer    | 2391 |
| Db                                   | 7141 | ATCATCTCTGTGAGAGCTCCCGATTCGCCCAATGGCCACCGCATCGGAACCTGTCT        | 7200 |
| Qy                                   | 2392 | ValTyrGlyValThrValIlePheSerCysAsnSerGlyTyrThrLeuValGlySerArg    | 2411 |
| Db                                   | 7201 | GTCTACGAGGACCAACAGCCATCTTCTCGCAATTCGGATACACACTGGTGGGCTCCAG      | 7260 |
| Qy                                   | 2412 | ValArgGluCysMetAlaAsnGlyLeuTrpSerGlySerGlyValArgCysLeu-----     | 2429 |
| Db                                   | 7261 | GTGCGTGAATGATGAGCCCAATGGCTGTGAGTGGCTCTGAAGTCCGCTGCTGCCACT       | 7320 |
| Qy                                   | 2429 | -----   | 2429 |
| Db                                   | 7321 | CAGACCAAGCTCACCTCATTTCTATAAGCTCTCTTGATGATGATGATGATGATGATGAT     | 7380 |
| Qy                                   | 2430 | -----AlaGlyHisCysGlyThrProGluProIleValAlaGlnGlyHisAlaAsnGly     | 2446 |
| Db                                   | 7381 | CTCACCAAGCTGAGACCTGTGGAGCTCTCGAGCCCATTTGTCACAGACATCAATGGG       | 7440 |
| Qy                                   | 2447 | GluAsnTyrSerTyrArgGlySerValValTyrGlnCysAsnAlaGlyPheArgLeuIle    | 2466 |
| Db                                   | 7441 | GAGAACTACAGCTACCGGGGAGTGTGTATCAACATGCAATGCGTCTCGGCTGATC         | 7500 |
| Qy                                   | 2467 | GlyMetSerValArgIleCysGlnGlnAspHisHisTrpSerGlyLysThrProPheCys    | 2486 |
| Db                                   | 7501 | GGCATGTCTGTGCGATCTGCCAGAGGATATCATGCTGTGGGAGAGACCTTCTGT          | 7560 |
| Qy                                   | 2487 | Val-----  | 2487 |
| Db                                   | 7561 | GTGCAATGTTAAGCAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG       | 7620 |
| Qy                                   | 2488 | -----ProIleThrCysGlyHisAspProGlyAsnProValAlaAsnGlyLeu           | 2501 |
| Db                                   | 7621 | GAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT     | 7680 |
| Qy                                   | 2502 | ThrGlnGlyAsnGlnPheAsnLeuAsnAspValValLysPheValCysAsnProGlyTyr    | 2521 |
| Db                                   | 7681 | ACTCAGGATTAACGATTTAATCTCAACGATGTGTCAAGTTGTTCGCAACCTGGGTAT       | 7740 |
| Qy                                   | 2522 | MetAlaGluGlyAlaAlaArgSerGlnCysLeuAlaSerGlyGlnTrpSerAspMetLeu    | 2541 |
| Db                                   | 7741 | ATGCTGTAGGGGCTGCTAGGTCCCAATGCTGCGCAGCGGCGAATGAGTGAATCTG         | 7800 |
| Qy                                   | 2542 | ProThrCysArgIleIleLeuAsnCysThrAspProGlyHisGlnGluAsnSerValArgGln | 2561 |
| Db                                   | 7801 | CCCACTGCAAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA     | 7860 |
| Qy                                   | 2562 | ValHisAlaSerGlyProHisArgPheSerPheGlyThrThrValSerTyrArgCysAsn    | 2581 |
| Db                                   | 7861 | GTCCAGCGCAGCGGCGCGGACAGGTTCACTTGGGACCACTGTGTTCACCGGTGAC         | 7920 |
| Qy                                   | 2582 | HisGlyPheTyrLeuLeuGlyThrProValLeuSerCysGlnGlyAspGlyThrTrpAsp    | 2601 |
| Db                                   | 7921 | CACGGCTTTCACCTCTCGGCGACCCAGTGTCTGAGTGCAGAGGAGATGAGCATGGGAC      | 7980 |
| Qy                                   | 2602 | ArgProArgProGlnCysLeu   | 2608 |
| Db                                   | 7981 | CGTCCCGCCCGCAGTGTCTC  | 8001 |
| RESULT 11                            |      |   |      |
| ADH72215                             |      |   |      |
| ID ADH72215 standard; DNA; 10989 BP. |      |   |      |
| XX                                   |      |   |      |
| AC ADH72215;                         |      |   |      |



```
Db 1075 ATCAGTGGCAGCGAAGCTCCATTCATATGCTAACTGCGATGACCTTCCCTCCAGTT 1134
Qy 16 ILeSerAsp-----CysCyalaSerAsnGlnArgHisSer 27
Db 1135 ATCAAGTAGCAAGAAATTGGCTAGACTCCATTTACCTCTGACAGCAACACCGACGCAA 1194
Qy 28 ValGlyValGlyProSerGluLeuValIleValysGlnIleGluLeuLysSerArgGlyVal 47
Db 1195 ---GGATTAAAGCTCAGTCCAGTCAAGTAAAGGCGGATTGAGTTGAATGCAAGAGGAGTTC 1251
Qy 48 LysLeuMetProSerLysAspAsnSerGlnLysThrSerValLeuThrGlnValGlyVal 67
Db 1252 AAGATGCTGCGCCAGCAGAGGATGAGCAATAAAACTCTGCTTTCAGCCAGAGAGTCTT 1311
Qy 68 SerGlnGlyHisAsnMetCysProAspProGlyIleProGluValArgGlyLysArgLeuGly 87
Db 1312 GCATTGCTCTCAGCATGTGCTAGATCTCTGGATTCCAGAAAATGGTAGAAGAGCAGGT 1371
Qy 88 SerAspPhe---ArgLeuGlySerSerValGlnPheThrCysAsnGlnGlyTyrAspLeu 106
Db 1372 TCCGACTTCAGTAGAGGTGGTGGCAAAATGATACGTTTATGTGAGACAAATTACGTCTC 1431
Qy 107 GlnGlySerLysArgIleThrCysMetLysValSerAspMetPheAlaIleTyrSerAsp 126
Db 1432 CAGGATCTTAAAGCATCAGCTGTCAGAGCTTACAGAGCAGCTCGCTGCTGGAGTGAC 1491
Qy 127 HisArgProValCysArgAlaArgMetCysAspAlaHisIleuArgGlyProSerGlyIle 146
Db 1492 CACAGGCCCATCTCGCAGCGCAGCAACATGTCATCTCGTGGAGCCACGAGGCGTCC 1551
Qy 147 IleThrSerProAsnPheProIleGlnTyrAspAsnAlaHisCysValTyrPheIle 166
Db 1552 ATTAACCTCCCTTAATTTATCCGTTTCAGATGACATGATATGACACATCTGTGTGGGTGATC 1611
Qy 167 ThrAlaLeuAsnProSerLysValIleLysLeuAlaPheGlnGluPheAspLeuArg 186
Db 1612 ACCACCAACCGACCGGACAAAGTCAATGATGCTTGAAGAAGTTTGAGCGAGGACA 1671
Qy 187 GlyTyrAspThrLeuThrValGlyAspGlyGlyGlnAspGlnLysThrValIleu 206
Db 1672 GGCATATACACCCCGACGGTGGTGTGATGCTGGAGAGTGGACACACAGATCGGTCTTG 1731
Qy 207 TyrMetSerGlnAsnAlaCysSerAspSerProHisThrProGlySerArgIleProGlu 226
Db 1732 TACGTG-----CTCAGCGGATCCAGTGTTCCTCGAC 1761
Qy 227 ---SerMetSerGlyAspIleTyrArgGlnLysThrValLeuGluIleCys 243
Db 1762 CTCATTGTGAGCATGAGCAACAGATGTGCTACAT----- 1797
Qy 244 ArgAspIleSerSerSerAspAlaArgSerGlySerValArgLysSerProLysThrSer 263
Db 1798 ---CTGCAGTCGATGATATAC----- 1815
Qy 264 AsnAlaValGluLeuValAlaProGlyIle-----ThrGluIleGlnGlnGly 278
Db 1816 ---ATGGCTCACCCTGCTTTAAAGCTGTTTACCAAGAAATTTAAAGGGA 1863
Qy 279 SerCysGlyAspProGlyIleProAlaTyrGlyArgArgGlnGlySerArgPheHis 298
Db 1864 GGGTGTGGGGAATCTCTGAATCCCGGCTATGGGAAGCGGACGGGACGAGTTCCTCAT 1923
Qy 299 GlyAspThrLeuLysPheGlnCysGlnProAlaPheGlnLeuValGlyGlnLysAlaIle 318
Db 1924 GAGAGATCACTACCTTTAAATGCCCCGGGCTTTGAGCTGTGGGGAGAGATTTATTC 1983
Qy 319 ThrCysGlnLysAsnAlaGlnIleTyrSerAlaLysLysProGlyCysValPheSerCysPhe 338
Db 1984 ACCGTGACAGACAAATCATGAGTGTGTCACAAACCCAGCTGTGTTTTCATGTTTC 2043
Qy 339 PheAsnPheThrSerProSerGlyValValLeuSerProAsnTyrProGluAspTyrGly 358
Db 2044 TTCAACTTTACGGCATCATCTGGATTAATCTGTCAACAAATTATCCAGAGAAATATGGG 2103
Qy 359 AsnHisIleuHisCysValTyrPheLysIleuAlaArgProGluSerArgIleHisIleuAla 378
Db 2104 AACACATGAACATGATGCTGTGTGATTAATCTCGGACCCAGGAAGTGAATTCACCTTAATTC 2163
Qy 379 PheAsnAspIleAspValGluProGlnPheAspPheLeuValIleLysAspGlyAlaThr 398
Db 2164 TTTAATGATTTGATGATTTGAGCTCAATTGACTTTCTCGCGGTCAAGATTAATGCAATT 2223
Qy 399 AlaGluAlaProValLeuGlnLysThrPheSerGlyAsnGlnLeuProSerSerIleThrSer 418
Db 2224 TCTGACATAACGTCTCTGGGTACTTTCTTGCAATGAAGTCCCTTCCAGCTGGCCAGC 2283
Qy 419 SerGlyHisValAlaArgLeuGluPheGlnThrAspHisSerThrGlyLysArgGlyPhe 438
Db 2284 AGTGGGCATATAGTTGCTGCTGATGAAATTTCAGTGTGACATTCATCTGCGCAGAGGTTTC 2343
Qy 439 AsnIleThrPheThrThrPheArgHisAsnGlnCysProAspProGlyValProValAsn 458
Db 2344 AACATCATTACACACACATTTGGTCAAGATGAGTCCATGATCTGGCATTCCTTAATAC 2403
Qy 459 GlyLysArgPheGlyAspSerLeuGlnLysThrSerIleSerPheLeuCysAspGlu 478
Db 2404 GGAACACGTTTGTGACAGCTTTTACTCGGAGAGCTGGTTTCTTTCACATGTGATGAT 2463
Qy 479 GlyPheLeuGlyTyrGlnGlySerGluThrIleThrCysValLeuLysGluGlySerVal 498
Db 2464 GGCCTTTGTCAAGACCAGGATCCGAGTCCATTCCTGATCATTGCAAGAGCGGAACCTG 2523
Qy 499 ValTyrAsnSerAlaValLeuArgCysGluAlaProCysGlyGlyHisIleuThrSerPro 518
Db 2524 GTCGTAGCTCACCGGTGCCCCGGTGTGAACCTCATGTGTGACATCTGCAGCGCTCC 2583
Qy 519 SerGlyThrIleLeuSerProGlyTyrProGlyPheTyrLysAspAlaLeuSerCysAla 538
Db 2584 AGCGGAGCATTTTGCTGCTCGATGGCCAGGAATATTAAGATTTCTTTCATTTGCA 2643
Qy 539 TyrValIleGlnAlaGlnProGlyTyrProIleLysIleThrPheAspArgPheLysThr 558
Db 2644 TGGATTAATGAAGAAACAGCGCAGCTCATCAAAATGACTTTTGACAGATTCACAGCA 2703
Qy 559 GluValAsnTyrAspThrLeuGluValArgAspGlyValArgThrLysSerAlaProLeuIle 578
Db 2704 GAGGTCAATTATGACACTGTTGAGAGTCAAGATGAGCCAGCCAGTTGCTCCACATGATC 2763
Qy 579 GlyValTyrHisGlyTyrGlnValProGlnPheLeuIleSerThrSerAsnTyrLeuTyr 598
Db 2764 GGCAGATACACCGCCACCGCAGGACCCAGTCTCATGACACCGGGAATTCATGTAC 2823
Qy 599 LeuLeuPheSerThrAspLysSerHisSerAspIleGlyPheGlnLeuAlaGlyTyrGluThr 618
Db 2824 CTGCTGTTCACACTGACCAACAGCCGCTCCAGCATGGCTTCTCATCACTATGAGAGT 2883
Qy 619 IleThrLeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGlnArgHis 638
Db 2884 GTGACGCTTGAATGAGATCTCTGCTGACCCGGGATCCCTGTGAACCGCATCCGCCAC 2943
Qy 639 GlyAsnAspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrThrLeu 658
Db 2944 GGTGAGACTTGTGACATGAGTCAACAGTGAATTTAGCTGTGACCCGGGGTACACACTA 3003
Qy 659 SerAspGlyLysProLeuGlnCysGluProAsnPheGlnTyrSerArgAlaLeuProSer 678
Db 3004 AGTGAGAGAGGACCTCTGCTGTGTGAGAGAACCACTGAAACCACTTGGCCAC 3063
Qy 679 CysGluAlaLeuCysGlyGlyPheIleGlnGlySerSerGlyTyrThrIleLeuSerProGly 698
Db 3064 TGGAGCGCTCTATGTGAGAGCTATCAAGAGGAAGATGGAACAGTCTTCTCTCGGG 3123
Qy 699 PheProAspPheTyrProAsnAsnLeuAsnCysThrTyrPheIleGluThrSerHisGly 718
Db 3124 TTTCCAGATTTTATTCAAACTCTTAACCTCAGGTGACCATTTGAAGTGTCTCATGGG 3183
```

QY 719 LysGlyVal1PhePheThrPheHisLeuGluSerGlyHisAspTyrLeuLeu 738  
DB 3184 AAGAGATTCAAAATATCTTTTCAACCTTTTCTTGAGAGTCCCAAGCATTTACTG 3243  
QY 739 IleThrGluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuPro 758  
DB 3244 ATCAGAGGAGTGAAGTATTTTCCAGAGCCGTTGCCAGCTCACCGGATCGGTGTGCT 3303  
QY 759 AlaProIleSerAlaGlyLeuThrGlyAsnPheThrAlaGlnValArgPheIleSerArg 778  
DB 3304 CATACGATCAAGCAGGAGCTGTTTGAAATCTTCACTGCCAGCTTCGGTTTATATCAAC 3363  
QY 779 PheSerMetSerTyrGlyGlyPheAsnIleThrPheSerGlyTyrAspLeuGluProCys 798  
DB 3364 TTCTCAATTTCTGTCGAGAGGCTTCAATATCATTTTCAAGATATGACTGGAGCATCT 3423  
QY 799 GluGluProGluValProAlaTyrSerIleArgGlyLeuGlnPheGlyValGlyAsp 818  
DB 3424 GATGATCTCGAGTCCCTGCTTCCAGCGAAGAAATGGTTTCACTTGGTGTGGAGAC 3483  
QY 819 ThrLeuThrPheSerCysPheProGlyTyrArgLeuGlyValThrAlaArgIleThrCys 838  
DB 3484 TCTCGAGGTTTCTCTGCTTCTCGGATATCGTTTGAAGGTGCCACCAAGCTTACCTGC 3543  
QY 839 LeuGlyGlyArgArgLeuTyrSerSerProLeuProArgCysValAlaGlyCysGly 858  
DB 3544 CTGGGTGGGGCCGCCGCTGTGTGAGTGCACCTCTGCCAAGGTGTGTGCCAATGTGCA 3603  
QY 859 AsnSerValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTyrAsn 878  
DB 3604 GCAAGTGTCAAGAGAAATGAAGAAATGATTAATGCTTCAAAATTTTCCATTCACATTAAT 3663  
QY 879 AsnAsnHisGlyCysIleTyrSerIleGlnThrGlnProGlyValGlyIleGlnLeuLys 898  
DB 3664 AATACCCATGAGTGTATCTATATAAATGAACAAGCCGCAAGGGATCACCTTATGA 3723  
QY 899 AlaArgAlaPheGluLeuSerGlyGlyValAspValLeuLysValTyrAspGlyAsnAsn 918  
DB 3724 ACAGAGACTTCCAGCTTTTGAGAGATCTCTAAGAGTATGATGAGAAAAGACAGT 3783  
QY 919 SerAlaArgLeuLeuGlyValPheSerHisSerGluMetGlyValThrLeuAsnSer 938  
DB 3784 TCCCTACGTCACATGGGCAAGTTCCTAAATGAATTCCTGGGGCTGATCTTAAACGC 3843  
QY 939 ThrSerSerSerLeuThrLeuAspPheIleThrAspAlaGluAsnThrSerGlyPhe 958  
DB 3844 ACATCCAAATCACTATGGCTAGATTCAACCAATGATCTGCACACCGAACCAAGTATT 3903  
QY 959 GluLeuHisPheSerSerPheGluLeuIleLysCysGluAspProGlyThrProLysPhe 978  
DB 3904 CAATCACTATACCAAGTTTGAATCTGTTAAATGTGAGATCCGGGATCCCTTAACTAC 3963  
QY 979 GlyTyrIlyValHisAspGlyGlyHisPheAlaGlySerSerValSerPheSerCysAsp 998  
DB 3964 GGCTATAGAGTCCGATGAAGGCACTTTCACGCACTGTATCTGTACAGTTGGAAC 4023  
QY 999 ProGlyTyrSerLeuArgGlySerGluLeuLeuCysLeuSerGlyGluArgThr 1018  
DB 4024 CCGGGGATACGCAATGATGAGCAACACCTGACCTGTTGAGTGGAGAGAGAGTGT 4083  
QY 1019 TyrAspArgProLeuProThrCysValAlaGlyCysGlyValThrValArgGlyGluVal 1038  
DB 4084 TGGAGCAAAACCACTACTTCTGTGATACGGAAATGTGTGTGATCATCATGACCCACA 4143  
QY 1039 SerGlyGlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIle 1058  
DB 4144 TCAGAGCAAAATATGTCTCCCTGGCTATCCAGCTCGATATGACAAACCTTCACTGACCC 4203  
QY 1059 TyrThrIleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThr 1078  
DB 4204 TGGATTATAGAGGAGACCCAGAAAGACCAATTAGCTTCAATTATGTTTGTGACAGC 4263

QY 1079 GluGluValHisAspValLeuArgIleThrAspGlyProValGluSerGlyValLeuLeu 1098  
DB 4264 GAGATGGTCAACATCTCTCAAGCTGTGGACCGGCGGTGGACAGTGAACATCTGCTG 4323  
QY 1099 LysGluLeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerVal 1118  
DB 4324 AAGAGTGAATGAGCTCCGCCCTTCCGGAGAGATCCACACACTTCACTCACTCAC 4383  
QY 1119 LeuGlnPheSerThrAspPhePheThrSerTyrGlnGlyPheAlaIleGlnPheSerVal 1138  
DB 4384 CTGAGATTGCACACAGCTTCTTCATCAGCAAGCTGTGCTTCTCATCTCACTTCTCAC 4443  
QY 1139 SerThrAlaThrSerCysAsnAspProGlyIleProGlnAsnGlySerArgSerGlyAsp 1158  
DB 4444 TCATTTGACGACCACTGTAAAGATTCAGTATGCCCCAATAATGCAACCCGTATAGAGAC 4503  
QY 1159 SerTyrGluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeuGlnGly 1178  
DB 4504 AGCAGAGAGGTGAGACACCGTCACATTCAGTGTGACCTGGCTATCAGCTCAAGGA 4563  
QY 1179 SerAlaGluIleSerCysValIlyValIleGluAsnArgPheThrGlnProSerPro 1198  
DB 4564 CAAGCCAAATCACCTGTGTGACAGTGAATTAACCGGTCTTTTGGCAACCAAGACCTCT 4623  
QY 1199 ThrCysIleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerPro 1218  
DB 4624 ACATTCATAGCTGCTGTGTGAGGAAATCTGACGGGCCCAAGAGGTATTTTGTACCC 4683  
QY 1219 AsnTyrProGluProTyrProProGlyLysGlyCysAspTyrIlyValThrValSerPro 1238  
DB 4684 AACTACCCACAGCCGATCTCTCGGGAAGAAATGTGATGAGAGATTAAGTAAGTAACCCG 4743  
QY 1239 AspTyrValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeu 1258  
DB 4744 GACTTGTCACTCGCTTGAATTAATCAAAAGTTTCAATGAGGCCACATGATGACTCTTA 4803  
QY 1259 HisIleTyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerGln 1278  
DB 4804 CACATCTATGAAGGGAGAGATTCCACAGCCCTCTCATTTGGAGTTTCCAGGGCTTCCAG 4863  
QY 1279 LeuProGlyValGlyLeuSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAla 1298  
DB 4864 GCCCAGAAAGAAATGAGAGATGAGGAAACAGCTGTTTCTGGCATTTCCGAGATGATGCC 4923  
QY 1299 SerValSerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCysPhe 1318  
DB 4924 TCCGTGGGCTTTCAGGGTTCCCATTTGAATTTAAAGAAACCAACCGGAAGCTTGTTT 4983  
QY 1319 AspProGlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeuGlySerSer 1338  
DB 4984 GACCCAGAAATATATGAAATGGAGCAAGAGTTGGAACAGACTTCAAGCTTGGCTTCAAC 5043  
QY 1339 ValThrTyrTyrCysHisGlyGlyTyrGluValGluGlyThrSerThrLeuSerCysIle 1358  
DB 5044 ATCAGCTACAGAGTGTACTGTGTTATGATTTCTGACCCCTCATCATCAGCTGTGTG 5103  
QY 1359 LeuGlyProAspGlyLysProValTyrAsnAsnProArgProValCysThrAlaProCys 1378  
DB 5104 ATTTGGGCTGATGGAAACCTCTCTGGGACCAAGTGTGCTTCTTCAATGTCTCCGT 5163  
QY 1379 GlyGlyGlnTyrValGlySerAspGlyValValLeuSerProAsnTyrProGlnAsnTyr 1398  
DB 5164 GAGGCGCAAGTACAGGATCAAGAGGGGTGATTTATCAACCAATCAACCCCATTAATAC 5223  
QY 1399 ThrSerGlyGlnIleCysLeuTyrPheValThrValProLysAspTyrValAlaPheGly 1418  
DB 5224 ACAGCTGTCAATATGCTCTTATTCATCAACCGTACCAAGAAATTCGTGTGCTTTGGA 5283  
QY 1419 GlnPheAlaPhePheHisThrAlaLeuAsnAspValValGluValHisAspGlyHisSer 1438  
DB 5284 CAGTTTGTCCATTTTCCAGACAGCCCTGAATGATTTGCAAAATTAATTTGATGGAACCAT 5343  
QY 1439 GlnHisSerSerLeuLeuSerSerLeuSerGlySerHisThrGlyGluSerLeuProLeu 1458



```
QY 2178 ILeaRgIYSerAlaProTYrCySeSerLeuProArGAlaProLeuHlaGlyPheIleu 2197
D 7564 ATTGGCTATGACGACCTTACTGCACTTGGACCCACCCCTGGAAGATGGGGGATTTCTA 7623
QY 2198 GlyInThrSerThrGlnProGlyGlySerIleHisPheGlyCySaAnaIaGlyTYrArg 2217
D 7624 AACAGACTGACGAGCGGTTGGAAAGCAAGTGCATTTTGGCAACCTGGATACCGA 7683
QY 2218 LeuValGlyHisSerMetAlaIleCySerThrArgHisProGlnGlyTYrHisLeuTrpSer 2237
D 7684 ATGGTCGGCCACAGCAATGCAACCTGTAGACGAACCACTGGCATGTACACAGTGGAC 7743
QY 2238 GluAlaIleProLeuCyGlnAlaLeuSerCySGlyLeuProGluAlaProLySaAnaIy 2257
D 7744 TCCCTCAAGCCACTCTGCCAGGCTGTCTCTGTGAAATCCCAAGATCCCGCAAGAACGCT 7803
QY 2258 MetValPheGlyLyGlyGlyTYrThrValGlyThrySaAnaIaTYrSerCySeSerGly 2277
D 7804 TCATTACCGGGAGAGAGTTCCTTGGACAGTAAAGTGTCTATGATGTCATGAAGGC 7863
QY 2278 TyThIleuGlnAlaGlyAlaGlyAlaThraIaGlyCyLeuAerPThrGlyLeuTrpSer 2297
D 7864 TTCAAGCTTGAAATCCAGCCAGCAACAGCCGTGTCTCAAGAAAGATGGGCTGTGGAGT 7923
QY 2298 AsnArgAenValProProGlnCyValProValThrCySProAerPValSerIleSer 2317
D 7924 AACAAAGGGAAGCCGCCCATGTGTAGCCGGTCCGTTCCCGACATGAAGCTACAGCTC 7983
QY 2318 ValGluHisGlyArgTrpArgLeuIlePheGluThrGlnTYrGlnPheGlnAlaGlnLeu 2337
D 7984 TCAGAACATGTCATCTGGAGGCTGTTCAAGATCTTGAATGAGTACGAGTCTCAAGTA 8043
QY 2338 MetLeuIleCyAerPProGlyTYrTYrTYrThrGlyGlnArgValIleArgCyGlnAla 2357
D 8044 TTGCTGAGCTCAGCTCTGTTACTACTTAGAAGCTGAGGCTCTCGCGGTGCGAGGCC 8103
QY 2358 AenGlyLySTrPserLeuGlyAAsPserThrProThrCySaArgIleIleSerCySGly 2377
D 8104 AATGGAGCTGGAAATGAGAGATGAGAGCCCAAGCTGCATGATTCGTGTGGAGGC 8163
QY 2378 LeuProIleProProAenGlyHisArgIleGlyThrLeuSerValTYrGlyAlaThraIa 2397
D 8164 CTTTCCTTCCCGCAATGGCAACAAGTTGGAACGTTGACAGATTATGGGGCCACAGCT 8223
QY 2398 IlePheSerCyAenSerGlyTYrThrLeuValGlySerArgValArgGlnCyMeCaIa 2417
D 8224 ATATTTACGTCAACACCGGCTACACGCTGTGGGCTCTCATGTACAGAGAGTGTGCA 8283
QY 2418 AenGlyLeuTrPserGlySerGlyValArgCySaLeuAlaGlyHisCySGlyThrPro 2437
D 8284 AATGGGCTCTGGAGGCGAGGAACCTGATGTTAGCTGGCCACTGGGTTCCCGCAGAC 8343
QY 2438 ProIleValAenGlyHisIleAenGlyGluAenTYrSerTYrArgGlySerValValTYr 2457
D 8344 CCGATTGTGAACGGTCATTAAGTGAAGATGGCTTCACTTACAGAGAACGGTGGTTAC 8403
QY 2458 GlnCySaAnaIaGlyPheArgLeuIleGlyMetSerValArgIleCySGlnGlnAerHis 2477
D 8404 CAGTGCATCTCGGTTTCCGGCTTGTGGAACTTCCGTGAGAGATATGCTTCCAGACAC 8463
QY 2478 HisTrPserGlyLySTrPserThrProPheCyValProIleThrCySGlyHisProGly 2497
D 8464 AAGTGGCTGGACAAACGCTGTCTGTGTCCGTATCACATGTGTGCACCCCTGGAAACCT 8523
QY 2498 ValAenGlyLeuThrGlnGlyAenGlnPheAenLeuAenAerValIaLYrPheValCyS 2517
D 8524 GCCACCGGATTCATTAATGGAGAGTTCACCGATGATGTGTGAATTTCACTGC 8583
QY 2518 AenProGlyTYrMetAlaGlyGlyAlaAlaArgSerGlnCyLeuAlaSerGlyGlnTrp 2537
D 8584 AACACGGGCTATTTGCTGACGGGCGTGTCTGAGCCAGTGTCCGAGCAACAGCGCACGTG 8643
QY 2538 SerAerMetLeuProThrCySaArgIleIleAenCySerThraerPProGlyHisGlnGlu 2557
D 8644 AGTAGCCCTGTGCCCAAGTGTGAGGTGAGTGAACCTGTCTGATCCAGCCTTGTGGAAAT 8703
QY 2558 SerValArgGlnValHisAaSerGlyProHisArgPheSerPheGlyThrThraIaSer 2577
D 8704 GCCATTGCTGACCGGCAACAGAACTCTCTGAGAGTTTGTGATGGAATGATATCTCTG 8763
QY 2578 TyTrArgCySaAnaIaGlyPheTYrLeuLeuGlyThrProValLeuSerCySGlnGly 2597
D 8764 TACCATTTGCAAGAGGATTTTACTGTGGATCTTCAAGCTTGAACCTGTATGGCAAT 8823
QY 2598 GlyThrTrPserAerPProAerProGlnCyLeuLeuValSerCySGlyHisProGlySer 2617
D 8824 GGCTTATGGAGCCATCTCCCTGCCAAGTGTGTGTGTATGTGTGTGACACCCAGGGGTC 8883
QY 2618 ProProHisSerGlnMetSerGlyAAsPserTYrThrValGlyAlaValaIaTYrSer 2637
D 8884 CCTGCACACGCGGCTCTCATGAGAGCTGTATTATGAGCGCGCTGTGCATCTAC 8943
QY 2638 CyGIleGlyLySaArgThrLeuValGlyAenSerThrArgMetCySGlyLeuAerGlyHis 2657
D 8944 TGCAGAGGAGCGAGACCTCATGAGCAACGACAGAGAGTGTCCAGAAAGACAGTCAAC 9003
QY 2658 TrPThrGlySerLeuProHisCySeSerGlyThrSerValGlyValCySGlyAAsPProGly 2677
D 9004 TGGAGCGGGGACCTGCCCTCATGAGCAAGAAATTAATCTGTGATTTGTGTGATCCGGGG 9063
QY 2678 IleProAlaHisGlyIleArgLeuGlyAAsPserPheAerProGlyThrValMetArgPhe 2697
D 9064 ACCCAGACATGGGCTCTCGGCTTGTGATGACTTTAAGCAAAAGATCTTCCGCTTC 9123
QY 2698 SerCySGlnAlaGlyHisValLeuArgGlySerSerGlnArgThrCySGlnAlaAenGly 2717
D 9124 TCTGTGAAATGGGGACAGCTGAGGGGCTCCCTTAAGCAGAGTGTGGTCAATGGG 9183
QY 2718 SerTrPserGlySerGlnProGluCyGlyValIleSerCySGlyAAsPProGlyThrPro 2737
D 9184 TCATGTGACAGACCTGACAGCCGCTGTGTGAGCCCTGTCTGTGGCAACCTGTGGACACC 9243
QY 2738 SerAenAlaArgValValPheSerAerPlyLeuValPheSerSerIleValTYrGlu 2757
D 9244 ACCAAACGAATGATGTGACAGATGAGGACTTGTCTTCCAGCTGGTCATGTATGCC 9303
QY 2758 CyAerGlyGlyTYrTYrAlaThrGlyLeuLeuSerArgHisCySaSerValAenGlyThr 2777
D 9304 TGCTGGAGAGGCTTAACAAGACCTCAGGGCTCATGACAGGCAATGGCACAGCCAAATGGACC 9363
QY 2778 TrPThrGlySerAerPProGluCyLeuValIleAenCySGlyAAsPProGlyIleProAla 2797
D 9364 TGGACAGGACCTGCTCCCGACTGCACAGTATTAAGTGTGGGATTCAGGCAACTAGCA 9423
QY 2798 AenGlyLeuArgLeuGlyAaAerPheArgTYrAenLySThrySaAnaIaTYrGlnCyVal 2817
D 9424 AATGACATCCAGTTTGGAGCAAGACTTCACTTCAACAAGACTGTGACATGATGATAC 9483
QY 2818 ProGlyTYrMetSerGlnSerHisArgValSerValLeuSerCyThrTYrLyAerArgThr 2837
D 9484 CCAGGCTATGTCAATGAGAGCACTACATCCGCACTATTAAGCTGTATACCAAGACGGCAG 9543
QY 2838 TrPserGlyThryLySTrPserProValCyLySaAlaLeuMetCyAlySTrProProLeuIlePro 2857
D 9544 TGGAAATCCAGAACCTGTGTCTGCAAGACCGTGTGTCTTCAAGCCGCGCGCTGACG 9603
QY 2858 AenGlyLySaValaGlySerAerPheMetTrPglySerSerValThrTYrAlaCyAlyLeu 2877
D 9604 AATGAAACAGTGGAGGAGAGATTTCCGGTGGGGCTCCAGCAATTAAGTTACAGTGCATG 9663
QY 2878 GlyGlyTYrGlnLeuSerLeuProAlaValaPheThrCySGlyGlyAenGlySerTrPThr 2897
D 9664 GAGGTTTACAGCTCTCTCACTCCGCGCATCTCTCTGTGAAGGTCCCGGGGTGTGAA 9723
QY 2898 GlyLeuLeuProGlnCyPheProValPheCyGlyAAsPProGlyValProSerArgGly 2917
```



```
Qy 322 LysAenAnGlnItrPSeSerAlaLeuYProGlyCyValPheSerCyPhePheAspPhe 341
Db 313 CAGAAACATCAGTGGTCTGGCAACAAGCCAGCTGTGATTTTCACTTTTCTTCACTTT 372
Qy 342 ThierProSerGlyValValLeuSerProAsnTYrProGluAspTYrGlyAsnHisLeu 361
Db 373 ACGGCATCATCTGGGATTAATCTGTGACCAATATATCAGAGAAATATGGAAACAACATG 432
Qy 362 HisCyValItrPLeuIleLeuAlaArgProGluSerArgIleHisLeuAlaPheAsnAsp 381
Db 433 AAGTGTCTGTGGATTAATCTCGAAGCCAGAAATCGAATTCACCTTAATCTTAATGAT 492
Qy 382 IleAspValGluProGlnPheAspPheLeuValIleLysAspGlyAlaThrAlaGluAla 401
Db 493 TTTGATGTGAGCTCAATTTGACTTTCTCCGGTCAAGATGATGATGATTTCTGACATA 552
Qy 402 ProValLeuGlyItrPheSerGlyAsnGlnLeuProSerSerIleThrSerSerGlyHis 421
Db 553 ACTGCTCGGTACTTTTCTGGCAATGAGTGCCTTCCAGCTGGCCAGAGTGGGAT 612
Qy 422 ValAlaArgLeuGluPheGlnItrPheAsnHisSerThrGlyValArgGlyPheAsnIleThr 441
Db 613 ATATGTGCTGTGAATTCAGTCACTTCCACTGACAGAGGATTCACATCACTACT 672
Qy 442 PheThrThrPheArgHisAsnGluCySerProAspProGlyValProValAsnGlyLysArg 461
Db 673 TACACCATTTGGTCAAGATGAGTGCATGATCTGGCATTTCTATTAAGAGAGAGT 732
Qy 462 PheGlyAspSerLeuGlnLeuGlySerSerIleSerPheLeuCyAspGluGlyPheLeu 481
Db 733 TTTGTGTCAGAGTTTCTACTCGGGAGTCTGGTTCTTTCCACTGATGATGATGCTTGTCTC 792
Qy 482 GlyItrGlnGlySerGluItrIleThrCyValIleLeuYsGlySerValItrPheAsn 501
Db 793 AAGCCCAAGGATCCGAGTCCATTAATCTGACTGCAAGAGCGGAGATGCTGTGAGC 852
Qy 502 SerAlaValLeuArgCyGluAlaProCySerGlyHisLeuThrSerProSerGlyItr 521
Db 853 TCCACCGTGGCCCTGTGAGTCCATGTGTGATGATCTGACAGGCTCCAGCGAGTCTC 912
Qy 522 IleLeuSerProGlyItrProGlyPheTYrLysAspAlaLeuSerCyAspAlaItrPheIle 541
Db 913 ATTTGCTCTCGATGGCCAGATATATTAAGATTTCTTAACATTTGATGATGATTAAT 972
Qy 542 GluAlaGlnProGlyItrProIleLysIleThrPheAspArgPheIleItrGluValAsn 561
Db 973 GAACCAAAACAGAGCCACTATCAAAATTAATTTTGAAGATTTGACAGAGTCAAT 1032
Qy 562 TyrAspThrLeuGluValArgAspGlyArgThrTYrSerAlaProLeuIleGlyValTYr 581
Db 1033 TATACACCTTGGAGTCAAGATGGCCAGCCAGTGTCTCCACATGATCGGAGATAC 1092
Qy 582 HisGlyItrGlnValProGlnPheLeuIleSerThrSerAsnTYrLeuTYrLeuLeuPhe 601
Db 1093 CACGGCAACCCAGGACCCAGTCTCATACAGACCCGGAACTTCAGTACTCTCTATTC 1152
Qy 602 SerThrAspLysSerHisSerAspIleGlyPheGlnLeuArgItrGluItrIleThrLeu 621
Db 1153 ACCACTACACAACACCGCTCCAGATCGGCTTCTCTCACTGAGAGTGAAGTCACTT 1212
Qy 622 GlnSerAspHisCyLeuAspProGlyIleProValAsnGlyAlaArgHisGlyAsnAsp 641
Db 1213 GAGTCGAGTTCCTCTCGAGCCGAGGATCCCTGTGAACGRCATTCGCAAGGAGAGAC 1272
Qy 642 PheTYrValGlyAlaLeuValItrPheSerCyAspSerGlyTYrItrLeuSerAspGly 661
Db 1273 TTTGGCATCAGGTCCACAGTGAATTTCAAGTGTGACCCGGGTACACACTAAGAGAGAC 1332
Qy 662 GluProLeuGluItrGluProAsnPheGlnItrPheArgAlaLeuProSerCyGluAla 681
Db 1333 GAGCCCTCTCTGTGAGAGAACACACAGTGAACAACACGCTTGGCCAGCTGTGAGAGCT 1392
Qy 682 LeuCyGlyGlyPheIleGlnGlySerSerGlyItrIleLeuSerProGlyPheProAsp 701
Db 1393 CTATGTGAGGCTATCATCCAGGAAAGATGGAACATCTCTTCTCGGTTCCAGAT 1452
Qy 702 PheTYrProAsnAsnLeuAsnCyThrTYrIleIleGlnItrSerHisGlyLysGlyVal 721
Db 1453 TTTTATCAAACTCTTAACACGCACTGAGCCATTAAGATGTCTCATGGGAAAGAGT 1512
Qy 722 PhePheThrPheHisItrPheHisLeuGlnItrSerGlyHisAspTYrLeuLeuIleThrGlu 741
Db 1513 CAATGATCTTTCACACCTTTCATCTTGAAAGTCCACAGACTAATTAATGATCAGACAG 1572
Qy 742 AsnGlySerPheThrGlnProLeuAsnGlnLeuThrGlySerArgLeuProAlaProIle 761
Db 1573 GATGAGATTTTTCGAGCCCGTGGCCAGGCTCACCGGGTCTGGTTTCTCATACATC 1632
Qy 762 SerAlaGlyLeuTYrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSerMet 781
Db 1633 AAGCAGAGCTGTGGAACCTTCACTGAGCCAGGTTGGTTTATACAGACTTCTCAAT 1692
Qy 782 SerTYrGluGlyPheAsnIleThrPheSerGluTYrAspLeuGluProCyGluGluPro 801
Db 1693 TCGTACAGAGGCTTCAATATCAATTTTCAGAAATATGATCAGAGCATGTGATGATCT 1752
Qy 802 GluValProAlaTYrSerIleArgLysGlyLeuGlnPheGlyValGlyAspThrLeuThr 821
Db 1753 GAGTCCCTCTCTACCCAGCCAGAAATGATGTTTCACTTGTGTGGAGACTCTCTAGC 1812
Qy 822 PheSerCyAspPheProGlyTYrArgLeuGluGlyItrAlaArgIleThrCyLeuGlyGly 841
Db 1813 TTTTCTGCTTCCGAGATATCGTTTGAAGGTGCCCAAGCTTACTCTCTCGGTGGG 1872
Qy 842 ArgArgArgLeuItrPheSerProLeuProArgCyValAlaGluCyGlyAsnSerVal 861
Db 1873 GCGCGCTGTGTGAGATGACCTCTGCAAGGTGTGTGCCAATGTGAGCAAGTCTC 1932
Qy 862 ThrItrGlnGlyItrLeuLeuSerProAsnPheProValAsnTYrAsnAsnAsnHis 881
Db 1933 AAAGAAATGAAGAACTATCTGTCTCCAAATTTTCCATCCAAITTTGATTAATACAT 1992
Qy 882 GluCyIleItrYserIleGlnItrGlnProGlyLysGlyIleGlnLeuYsAlaArgAla 901
Db 1993 GAGGTATCTATAAATATGAAGACGAAGCCGAGGATTCACCTTAAGACAGAGC 2052
Qy 902 PheGluLeuSerGluGlyAspValLeuLysValItrAspGlyAsnAsnSerAlaArg 921
Db 2053 TTCACGTGTGTAAGAGATCTATTAAGGTATATGATGAAGAAGCACTTCTCAGT 2112
Qy 922 LeuLeuGlyValPheSerHisSerGluMetMetGlyValItrLeuAsnSerThrSerSer 941
Db 2113 CCACTGGGACGTTCACTATAAAATGAATCTTGGGCTGATCTTAACACAGCATCCAA 2172
Qy 942 SerLeuItrPheAspPheIleThrAspAlaGluAsnItrSerLysGlyPheGluLeuHis 961
Db 2173 CACTTGTGCTAGGTTCAACACCAATGATCTGACACCGACCAAGGTTTTCAACTCAC 2232
Qy 962 PheSerSerPheGluLeuIleLysCyGluAspProGlyItrProLysPheGlyTYrLys 981
Db 2233 TATACAGTTTATATCTGTAATATGAGATCCGGGCAATCCCTAATCAGGCTAATAG 2292
Qy 982 ValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCyAspProGlyTYr 1001
Db 2293 ATCGATATGAAGCCACTTACCGACACTGTGATTTGTACAGTTTCAACCCGGGATAC 2352
Qy 1002 SerLeuArgGlySerGluGluLeuLeuCyLeuSerGlyGluAlaArgItrTYrAspArg 1021
Db 2353 GCCATGATGAGCAACAACCTGACCTGTTTATGAGGAGACAGAGAGTGTGGACAAA 2412
Qy 1022 ProLeuProThrCyValAlaGluCyGlyGlyItrValAlaArgGlyValSerGlyGln 1041
Db 2413 CCACTACTTTCGTCATGAGGATGTGGTGTGATTCATTCAGCCACATCAGAGCA 2472
Qy 1042 ValLeuSerProGlyTYrProAlaProTYrGluHisAsnLeuAsnCyIleItrPheIle 1061
```

```

Db      2473  ATATTGTCCTCCGCTATCCAGCTCCGATGACAAACAACCTCAGCTGACCTGATTATA 2532
Qy      1062  GIUALAGLUALAGLYCyethr11leglyleuH1sPheLeuVal1PheAspThrGluGluVal 1081
      2533  GAGGAGAACCCGAGAAAGACATTAGCTTCCATTTCATTTGTCAGACGAGATGGCT 2592
Qy      1082  HisAspVal1LeuArg11eTyrAspGlyProVal1GluSerGlyVal1LeuLeuGly1Leu 1101
      2593  CACGACATCTCAAGAGTCTGGGACGGGCGGTGGACATGACATCTCTGCTAAGAGATGG 2652
Qy      1102  SerGlyProAlaLeuProValAspLeuH1sSerThrPheAsnSerVal1ValLeuGlnPhe 1121
      2653  AGTGCTCCGCTTCGAGAGACATCCAGACACCTTCACTCACTCACTCACTCACTTC 2712
Qy      1122  SerThrAspPhePheThrSer1ySerGln1yPheAla11eGlnPheSerVal1SerThrAla 1141
      2713  GACGCGACCTTCTCATACGACAGCTGGCTTCTCCATCTCCAGTTCCTCACTCAATTGCA 2772
Qy      1142  ThrSerCyAsnAspProGly11eProGlnAsn1ySerArgSerGlyAspSerTrpGlu 1161
      2773  GCCACCTGTAAAGATCCAGATGAGCCCAAAATGGACCCGCTATGGAGACAGAGAG 2832
Qy      1162  AlaGlyAspSerThrVal1PheGlnCyAspProGly1yTrAla1eGln1ySerAlaGlu 1181
      2833  GCTGGAGACACCTGCACTTCACAGTGTACCTGCTATCACTCAAGAGCAAGCCAA 2892
Qy      1182  11eSerCyVal1y11eGlnAsnArgPhePheTrpGlnProSerProProThrCys11e 1201
      2893  ATCACTGTGTGACGATTAACCGGTCTCTTGGACACAGACCTCTCACTGATGATA 2952
Qy      1202  AlaProcyAspGly1yAspLeuThrGly1yProSerGlyVal11eLeuSerProAsnTrpPro 1221
      2953  GCTCTGTGTGAGGAAATCTGACGGGCCAGCAGGTGTATTTGTACCACTCACTCA 3012
Qy      1222  GluProTrpProProGly1yGluCyAspTrpVal1ThrVal1SerProAspTrpVal 1241
      3013  CAGCGATATCTCTCTGGGAGAGATGTGACTGAGATTAAGTAAACCCGACCTTGTTC 3072
Qy      1242  11eAlaLeuVal1PheAsn11ePheAsnLeuGlnProGly1yTrAspPheLeuH1s1eTrp 1261
      3073  ATCCCTTGATATTAACAAGTTTCAACATGAGCCGATGATGCTTCCATACATCTAT 3132
Qy      1262  AspGlyArgAspSerLeuSerProLeu11eGlySerPheTrpGlySerGlnLeuProGly 1281
      3133  GAGGAGGAGATTCACACAGCCCTCTATGGAATTAACAGAGGCTCTCAGGCCCAAGA 3192
Qy      1282  Arg11eGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerValSer 1301
      3193  AGAATAGAGATGACGGAACAGCTCTTCTGCACTTTCGAGTATGCTCCGTGGCC 3252
Qy      1302  AsnAlaGlyPheVal11eAspTrpThrGluAsnProArgGluSerCyAspPheAspProGly 1321
      3253  CTTCAGGCTTCGCTCATTTGAATTTAAAGAGAAACAACGAGGACCTTGTTTGAACCA 3312
Qy      1322  Ser11eLeuAsnGlyThrArgVal1GlySerAspLeuLeuGln1ySerSerVal1ThrTrp 1341
      3313  AATATATGATGATGAGACAGAGATTTGAACAGACTTCAAGCTTGCTCCACATCACTAC 3372
Qy      1342  TyrCyHis1eGly1yTrpGluVal1GluGlyThrSerThrLeuSerCys11eLeuGlyPro 1361
      3373  CAGTGTGACTCTGGCTATTAAGATTCTTGACCCCTCATTCATCACTGCTGATGGGGCT 3432
Qy      1362  AspGlyLysProVal1TrpAsnAsnProArgProVal1CysThrAlaProCysGlyGln 1381
      3433  GATGGAGAAACCTTCGAGGACCAAGTGTCTCTCTGCAATGCTCCCTGTGAGAGCCAG 3492
Qy      1382  TyrVal1GlySerAspGlyVal1Val1LeuSerProAsnTrpProGlnAsnTrpThrSerGly 1401
      3493  TACACGGATCAGAAAGGATGATTTATACCAAACTTACCCCATATATACCACTGAT 3552
Qy      1402  Gln11eCysLeuTrpPheVal1ThrVal1ProLysAspTrpVal1Val1PheGlyGlnPheAla 1421
      3553  CAATATGCTCTATATCCATCAAGGATCAACAAAGAAATGTGTCTTGGACAGATTGCC 3612
      Db

Qy      1422  PhePheHisThrAlaLeuAsnAspVal1Val1GluVal1HisAspGlyH1sSerGlnH1sSer 1441
      3613  TATTTCAGACAGCCCTGATGATTTGGACAGATTTTGAATGGAACCATGACAGAGCC 3672
Qy      1442  ArgLeuLeuSerSerSerSerGlySerH1sTrpGly1ySerLeuProLeuAlaThrSer 1461
      3673  AGACTTCACTGACTCTCTGGGGTCTCACTAGGAGGAAACATGCTCTGTGCTATGCTCA 3732
Qy      1462  AsnGlnVal1Leu11eLysPheSerAla1yAspGlyLeuAlaProAlaArgGlyPheHisPhe 1481
      3733  AATCAATCTGCTCCGATTCAGTGCAAGAGCGGGCTCTGCGCGGCTTCCACTTC 3792
Qy      1482  ValTrpGlnAlaVal1ProArgThrSerAla1ThrGlnCysSerSerVal1ProGluProArg 1501
      3793  GTGTATCAAGCTGTTCCTCTGTACAGTGACACCCATGACGCTCTGCTCCAGCCAGA 3852
Qy      1502  TyrGlyLysArgLeuGly1ySerAspPheSerVal1GlyVal11eVal1ArgPheGlyCysAsn 1521
      3853  TACGAGAGAGATTTGTTCTGATGATTTCTGCGGCTCATGCTCGATTCAGATRCAC 3912
Qy      1522  SerGlyTrpAlaLeuGln1ySerProGlu11eGlyCysLeuProVal1ProGlyAlaLeu 1541
      3913  CCGGATACCTGCTCTCAGGGTCCACGGGCTCCAGTGCAGTCCGTGCGCAAGCCTTG 3972
Qy      1542  AlaGlnTrpAsnVal1SerAlaProThrCysVal1Val1ProCysGlyGlyAsnLeuThrGlu 1561
      3973  GACAGTGGAGACACATCCCACTGTGTGTATCCCTGACAGTGGCAATTTCACTCA 4032
Qy      1562  ArgArgGlyThr11eLeuSerProGlyPheProGluProTrpTrpLeuAsnSerLeuAsnCy 1581
      4033  CGAAGGATGACATCTGCTCCCGGCTACCTGAGCCATACGAAACAATTTGAACCTGT 4092
Qy      1582  ValTrpLys11eVal1Val1ProGluVal1AspGly11eGln11eGlnVal1SerPheVal 1601
      4093  AATATGAGATCATGTATACGAGAGGCTCGGGAATTCAGATCAAGTATGATGATTTGCC 4152
Qy      1602  ThrGlnGlnAsnTrpAspSerLeuGlnVal1PheAspGlyVal1AspAsnThrVal1ThrMet 1621
      4153  ACGAGAGCAATCTGGAATCTCTTGAATTCACAGATGTGGAGATGTGACGCCACCA 4212
Qy      1622  LeuGlySerPheSerGlyThrThrVal1ProAlaLeuAsnSerThrSerAsnGlnLeu 1641
      4213  CTGGGAACCTTCTGACGACACAGTACCGGCACTGCTGAGACATCTTCAACCACTC 4272
Qy      1642  TyrLeuHisPheTrpSerAsp11eSerVal1SerAla1eGlyPheH1sLeuGlnTrpLys 1661
      4273  TACCTGCAATTCAGTCTGACATTAAGTGTGACAGCTGTGTTTCCACCTGGAATACAA 4332
Qy      1662  ThrVal1GlyLeuSerSerCyProGluProAlaVal1ProSerAsnGlyVal1Val1ThrGly 1681
      4333  ACTGATGATCTTCTGCTGATGCAAGAACAGACCCCTCCCAAGACATCAAAATGGGA 4392
Qy      1682  GluArgTrpLeuVal1AsnAspVal1Val1SerPheGlnCysGluProGly1yTrAlaLeuGln 1701
      4393  GATCGATACAGTGAACAGAGTGTCTCTCTTCAAGTCGAGCCGGGTACACCTTGAG 4452
Qy      1702  GlyHisAlaHis11eSerCysMetProGly1yThrVal1ArgArgTrpAsnTrpProProPro 1721
      4453  GGCGGTTCCACATTTCTGTATGCCAGGAGACGTTTGGCTGGAACATTCCTCTGCC 4512
Qy      1722  LeuCy11eAlaGlnCysGlyGlyThrVal1GluGluMetGlyVal111eLeuSerPro 1741
      4513  CTGTGATTTGAACCTGTGAGAGGACCTAGACACTTGGGTGTGATCTCGAGGCCCC 4572
Qy      1742  GlyPheProGlyAsnTrpProSerAsnMetAspCysSerTrpLys11eAlaLeuProVal 1761
      4573  GGCTTCCAGATTTCTTACCCCAACTTATGATGACCTGAGAGATCTCATTAACCATC 4632
Qy      1762  GlyPheGlyVal1His11eGlnPheLeuAsnAspSerThrGluProAsnHisAspTrp11e 1781
      4633  GGCTATGTGACATATTCAGTTTCTGAATTTTTCACGAACTATATCATGACTTCCCT 4692
      Db
```

```
Qy 1782 Glu1IeaRgaNGlyProTyRGluThrSerArgMetSerGlyArgPheSerGlySerGlu 1801
Db 4693 GAAATTCAAAATGGAACCTTACCAACAGCCCATGATGGACAATTTAGGGGACCGAT 4752
Qy 1802 LeuProSerSerLeuLeuSerThrSerHisGluThrValTyrPheHisSerAspHis 1821
Db 4753 CTCCCGCGCGCCCTGCTGACGACAAACGATGAACCCCTACCTCACTTTATAGTACCAT 4812
Qy 1822 SerGlnPheArgProGlyPheLeuLeuGluTyrGlnAlaTyrGluLeuGlnGluCyPro 1841
Db 4813 TCGGAAAAACCGGACAGATTAACTTGCTTACCAAGCCATAGATTGACAAACCTGCCA 4872
Qy 1842 AspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTyrAsnValGlyGlnSer 1861
Db 4873 GATCAACCCCATTTCCAGAAATGGGTACATGATCACTCGGATTCACAGGTGGGCAATCA 4932
Qy 1862 ValThrPheGluCyLeuProGlyTyrGlnLeuThrGlyHisProValLeuThrCySerGln 1881
Db 4933 GTATCTTTCAGAGTTTATCTGGGTACATCTTAATAGGCCATCCGTCTCATCTTGTCAG 4992
Qy 1882 HisGlyThrAsnArgAsnTyrAspHisProLeuProGlySerGlyValProCyGlyGly 1901
Db 4993 CATGGGATTCACAAACAACTGGAACCTACCTTTCCAAAGATGATGCCCTTGGGGTAC 5052
Qy 1902 AsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyrSerSer 1921
Db 5053 AACGTAACTTCTCGAAGCGACCACTACTCTCCCTGCTTCCCTGATGATATCCGATC 5112
Qy 1922 SerGlnAspCyValTyrLeuIleThrValProIleGlyHisGlyValArgLeuAsnLeu 1941
Db 5113 CTGAAGACCTGATTTGGCTCATCAGCGTGCCTCAGGCGACGAGATTACATCAACTTC 5172
Qy 1942 SerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTyrAspGlyProGlnGln 1961
Db 5173 AACCTGTATACAGACGAGAGCTGTCAACGATTAATCTGTGGAGCGGCTCCGATCAG 5232
Qy 1962 ThrAlaProArgLeuGlyValPheThrArgSerMetAlaIleValThrValGlnSerSer 1981
Db 5233 AACTCACCACCGCTGGAGTTTTCAGTGGCAACAGCCCTCGAAACGGCGTATAGCTCC 5292
Qy 1982 SerAsnGlnValLeuLeuLysPheHisArgAspAlaAlaThrGlyGlyIlePheAlaIle 2001
Db 5293 ACCAACCAAGCTCTGCTCAAGTTCACAGCATTTTCAATGAGAGCTTCTTGTCCTC 5352
Qy 2002 AlaPheSerAlaTyrProLeuThrLysCyProProProProThrIleLeuProAsnAlaGlu 2021
Db 5353 AATTTCACGCAATTCAGCTCAGAAATGTCAACCTCCCGAGCGGTTCCACAGCGAGAA 5412
Qy 2022 ValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrArgCyLeuPro 2041
Db 5413 ATGCTTACTGAGATGATGATTTGAGATAGAGATTTTGTGAAGTACCAAGTGCACCCC 5472
Qy 2042 GlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuGlyThrTyrLeuGlnPhe 2061
Db 5473 GGGATACACCTTGGTGGGACCGACATCTGACTTGGAACTCACTCCCACTTGCAGATT 5532
Qy 2062 GluGlyProProProIleCyGlyValHisCyProThrAsnGluLeuLeuThrAspSer 2081
Db 5533 GAGGGTTCTCTCCCAACATGTGAAGCAACATAGCCAGCAATGAAAGTCCGACGTGATCA 5592
Qy 2082 ThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCySer 2101
Db 5593 TCGGAGATTCATCGACCGAGGATCCGGGTAAATTTTAACTCCCAAGACTTGCTCT 5652
Qy 2102 TyrLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeuSer 2121
Db 5653 TGGAGTATTAAGTGAACCAACTACAACTTACATCTTGTGGACACATTTCAAGT 5712
Qy 2122 GluLysGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeuLeu 2141
Db 5713 GAAAAGCAGTTTGTATGCACTGGAAGTGTGTGATGTTCTTCTGGGCAAAAGTCTCTCTA 5772
Qy 2142 LysAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerSerAsnSerVal 2161
Db 5773 GTAGCTTAAGTGGGATATCACTAGCAACATTTTACAAAGCAGAGATTAACATTA 5832
Qy 2162 TyrLeuArgTyrPheSerAspHisAlaTyrAsnArgLysGlyPheLysIleArgTyrSer 2181
Db 5833 TATCTCCGCTGGTCCATGACATGCATGCCACCAAGAAAGAAAGATTCAGATTCGTATGCA 5892
Qy 2182 AlaProTyrCySerSerLeuProArgAlaProLeuHisGlyPheIleLeuGlnThrSer 2201
Db 5893 GCACCTTACTGCAAGTTTGACCCACCCCTGAGAAATGGGGTATTTTAAACGAGCTGCA 5952
Qy 2202 ThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGlyHis 2221
Db 5953 GAGCGGTGGAGCAAAAGTCATTTATTTTGCAGCGCTGGATACCGAATGTCGGGCAC 6012
Qy 2222 SerMetAlaIleCySerThrArgHisProGlnGlyTyrHisLeuTyrSerGluAlaIlePro 2241
Db 6013 AGCAATCAACCTGTAGACAAACCCACTTGGCATGACAGTGGGCACTCCCTACCGCA 6072
Qy 2242 LeuCyGlnAlaLeuSerCySerGlyLeuProGluAlaProLysAsnGlyMetValPheGly 2261
Db 6073 CTGTCCAGGCTGTGCTCTGTGATCCAGAAATCCCAAGAAACGGTTCAATTTACCGGG 6132
Qy 2262 LysGluTyrThrValGlyThrLysAlaValTyrSerCysSerGluGlyTyrHisLeuGln 2281
Db 6133 AACGATTCATTTGGACAGTAAAGTGTCTATGAAATGTCATAGAGGCTTCAAGCTTGA 6192
Qy 2282 AlaGlyAlaGluAlaThrAlaGluCyLeuAspThrGlyLeuTyrSerAsnArgAsnVal 2301
Db 6193 TCCAGCCAGCAAGCAACAGCCGTGTGTCAGAAAGATGGCTGTGGAGTTAACAGGGGAG 6252
Qy 2302 ProProGlnCyValProValThrCyProAspValSerSerIleSerValGluHisGly 2321
Db 6253 CCGCCACGGTAAAGCCGGTGGCTTGGCCACGATTAACCTCACTCAGAAACATGTC 6312
Qy 2322 ArgTyrArgLeuIlePheGluThrGlnTyrGlnPheGlnAlaGlnIleuMetLeuIleCy 2341
Db 6313 ATCTGAGAGCTGTTCAGATCTCTGAAAGATGATAGCTCTCAAGTATGCTGAGCTGC 6372
Qy 2342 AspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLysTyr 2361
Db 6373 AGTCTGTGTACTTACCTTAGAAGGCTGGAGCTGCTCGCGGTCAGAGCCAAATGGGACGTGG 6432
Qy 2362 SerLeuGlyAspSerThrProThrCyArgIleIleSerCyGlyGluLeuProIlePro 2381
Db 6433 AACATAGAAATAGAGAGCCCAAGCTGTGAGTATCTGCTGTGAAACCTTCTCTTCC 6492
Qy 2382 ProAsnGlyHisArgIleGlyThrLeuSerValTyrGlyValAlaThrAlaIlePheSerCy 2401
Db 6493 CCAATAGCAACAAAGATTGAAAGCTTACAGATTTATGGGCCACAGCTATATTTAGCTGC 6552
Qy 2402 AsnSerGlyTyrThrLeuValGlySerArgValArgGluCysMetAlaAsnGlyLeuTyr 2421
Db 6553 AACACCGGCTAACAGCTGTGGGGTCTCATGTCAAGAGTGTGGCAATATGGGCTGTGG 6612
Qy 2422 SerGlySerGlyValArgCysLeuAlaGlyHisCyGlyThrProGluProIleValAsn 2441
Db 6613 AGCGGACGCAAACTCATCTGCGCGGCTGCGGCTGCGGCTCCCAAGCCGATTTGTAAAC 6672
Qy 2442 GlyHisIleAsnGlyLeuAsnTyrSerTyrArgGlySerValValTyrGlnCysAsnAla 2461
Db 6673 GGTCACTTATGTGAGATGAGCTTCAAGTACAGACACGGTGTATTCAGCTCAATCTCT 6732
Qy 2462 GlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisIleTyrSerGly 2481
Db 6733 GATTTCGGGCTTGGGAACTTCCGTGAGATATGCTCGCAAAACCAAGTGTGTGCA 6792
Qy 2482 LysThrProPheCySerValProIleThrCyGlyHisArgGlyAsnProValAsnGlyLeu 2501
Db 6793 CAAACGCTCTGTGTGCTCCCATCACTGAGTGCACCTGTGAAACCTTGCCACGAGATTC 6852
Qy 2502 ThrGlnIleAsnGlnPheAsnLeuAsnArgValValLysPheValCyAsnProGlyTyr 2521
```

```

Db 6853 ACTAATGGCAGTAGTTCACCTGAATGATGTCGTAATTTCACCTGCAACAGGGGCTAT 6912
Qy 2552 MetAlaGluGlyValAlaAlaArgSerGlnGlyLeuAlaSerGlyGlnTTPSerAspMetLeu 2541
Db 6913 TTGCTGCAAGGCGGTCTCTGAGCCAGTGTGAGCAACGGCCAGTGAAGAGCCCTG 6972
Qy 2542 ProThrCybArG1le1leAsnCybThrAspProGlyNH1asGlnIuAsnSerValArgGln 2561
Db 6973 CCCAGTGTGAGAGGTGTAAGTCTTCTGATCCAGGCTTTGAGAAAATGCCATTCTGCAC 7032
Qy 2552 ValHsAlaSerGlyProHsArgPheSerPheGlyThrThrValSerTyArgCybAsn 2581
Db 7033 GGGCAACAACACTTCCCTGAGAGATTTCAGATGAGTAATGAGTATCTGTACATTCAG 7092
Qy 2582 HisGlyPheTyLeuLeuGlyThrProValLeuSerCybGlnGlyAspGlyThrTPAsp 2601
Db 7093 AAGGATTTTCTTCTGCGGATCTTCAGCCTTGACCTGTATGCGCAAAATGGCTTATGGAC 7152
Qy 2602 ArgProArgProGlnCybLeuLeuValSerCybGlyNH1asProGlySerProProHsSer 2621
Db 7153 CGATCCCTGCGCCAGTAGTGTGGCTATATCGTGTGACACCCAGGGGCTCCCTGCCAAGCC 7212
Qy 2622 GlnMetSerGlyAspSerTyThrValGlyAlaValAlaArgTySerCyb1leGlyLys 2641
Db 7213 GTCTCACTGAGAGCTGTTTACCTATGCGCGCTCGTGCATATCTCTGACAGAGGAGC 7272
Qy 2642 ArgThrLeuValGlyAsnSerThrArgMetCybGlyLeuAspGlyNH1asTPThrArgLys 2661
Db 7273 GAGAGCTCATAGGCAACGACACGAGGTGTCCAGGAAAGACGTCACTGAGACGGGAGC 7332
Qy 2662 LeuProHsCybSerGlyThrSerValGlyValCybGlyAspProGlyLysProAlaHs 2681
Db 7333 CTGCCCACTGCACAGAGAAATTAATCTGATCTGTGTATCGGGAGCCCAAGACAT 7392
Qy 2682 GlyLeuArgLeuGlyAspSerPheAspProGlyThrValMetArgPheSerCybGlyAla 2701
Db 7393 GGGTCTGGCTGTGATGATCTTAAACAACAAGAGCTTCTCGCTCTCTCTGTGTAAGT 7452
Qy 2702 GlyHsValLeuArgGlySerSerGlyArgThrCybGlnAlaAsnGlySerTPSerGly 2721
Db 7453 GGGGACCAAGCTGAGGGGCTCCCTCGAAGCAGCATGTTGTCTCAATGGTTCATGTGAG 7512
Qy 2722 SerGlnProGlyCybGlyVal1leSerCybGlyAsnProGlyThrProSerAsnAlaArg 2741
Db 7513 CTGACAGCGGTGTGTGAGCGGTCTCTGTGCGCAACCTGTGCACACCAACGGAATG 7572
Qy 2742 ValValPheSerAspGlyLeuValPheSerSer1leValTyArgGlyAspArgGly 2761
Db 7573 ATTGTCAGTAGTATGATTCATTTCTCTCAAGCTCGGTCACTATGCTGTGGAGAGC 7632
Qy 2762 TyThrAlaThrGlyLeuLeuSerArgHsCybSerValAsnGlyThrTPThrArgLys 2781
Db 7633 TACAAAGACCTCAGGGCTCATGACAGGCAATTGACAGCCATGGGACCTGACAGGCACT 7692
Qy 2782 AspProGlyLeuVal1leAsnCybGlyAspProGlyLysProAlaAsnGlyLeuArg 2801
Db 7693 GCTCCGACCTGCAATTAATGTTGTGGAGTCCAGGACCACTGAGCAATGGCATTCAG 7752
Qy 2802 LeuGlyAsnAspPheArgTyArgLeuValThrValThrGlyGlnCybValProGlyTyMet 2821
Db 7753 TTTGGGACGACCTTCACTTCAACAGACTGAGCTGATCAAGTATCAACAGGCTATGTC 7812
Qy 2822 MetGlnSerHsArgValSerValLeuSerCybThrLysAspArgThrTPAsnGlyThr 2841
Db 7813 ATGGAGACGATCATCCACTATTCGCTGTACCAAGAGCGGAGCGGAGTATCCGAGC 7872
Qy 2842 LysProValCybValAlaLeuMetCybLysProProProLeu1leProAsnGlyLysVal 2861
Db 7873 AAACCTGTGTGCAAAACCGGTGTGTCTTCAGCCCGCCGCTGACAGATGGAACAGTG 7932
Qy 2862 ValGlySerAspPheMetTPArgLysSerValThrTyArgAlaCybLeuGlnGlyTyArg 2881
Db 7933 GAGGAGAGTATTTCCGCTGGGCTCCACATTAAGTATGACGTGATGAGCGGTTACCG 7992

```

---

```

Qy 2882 LeuSerLeuProAlaValPheThrCybGlnGlyAsnGlySerTPThrGlyLeuLeuPro 2901
Db 7993 CTCTCTCACTCGCGCCACTCTCTCTGTGAAGGTGCGGGGTGTGAAAGAGATCCCC 8052
Qy 2902 GlnCybPheProValPheCybGlyAspProGlyValProSerArgGlyValArgGluAsp 2921
Db 8053 CAGTGTCTCTGTGTGTGCGGAGACCTGTGCATCCCGCAGAAAGGGGACTTATGTGG 8112
Qy 2922 ArgGlyPheSerTyArgSerSerValSerPheSerCybHsAspProLeuValLeuVal 2941
Db 8113 AAAAGTTTCACTATTAAGTCCGAAGTCTTCTTCCAGTGAATCTCCATTTATCTCG 8172
Qy 2942 GlySerProArgArgPheCybGlnSerAspGlyThrTPSerGlyThrGlnProSerCy 2961
Db 8173 GGATCTCCAGAAAGTGTGCGCAACGTACAGGCACTGAGGCGGATCAACCCACCTGC 8232
Qy 2962 IleAspProThrLeuThrThrCybAlaAspProGlyValProGlnPheGlyLysGln 2981
Db 8233 ATTTGATCTGTCTATTAACACCTGCGCAACCTGTGTACGCCACACTTGGATTAACA 8292
Qy 2982 AsnSerGlnGlyTyArgValGlySerThrValLeuPheArgCybGlnLysGlyTyArg 3001
Db 8293 AGCTCAAGAGGCTATGAGGTGGAAGCAAGGTTTTTTCAGGTGACGAAAGCTTACAT 8352
Qy 3002 LeuGlnGlySerThrThrArgThrCybLeuProAsnLeuThrTPSerGlyThrProPro 3021
Db 8353 ATTCAGAGTTCACACATCGACCTGCTTCCATTTATGATGAGTGGATACAGACC 8412
Qy 3022 AspCybValProHsHsCybArgGlnProGlyNH1asProHsAlaAsnValGlyAla 3041
Db 8413 GAATTTATACCTCATAGCTGCGACAGCAGCAAAACCCCGGACACGCGAGTGTGAGACC 8472
Qy 3042 LeuAspLeuProSerMetGlyTyThrLeuLeu-ThrProAlaArgAlaSerProSe 3061
Db 8473 ATCGATCTTCTCTACTTGTGGCTACACCTTATGTATACCTGCATCCAGGCTTTTCTC 8532
Qy 3061 PArgValAlaProSerThrAlaProAlaArgMetAlaAlaGlyGlnAlaSerArg 3080
Db 8533 GCAAGGAGATCTGAGCACAAGAACTGTAACAGACATGAATATGACAGAAAGTCCG 8590

```

RESULT 13  
AAD33318  
ID AAD33318 standard; cDNA; 10673 BP.  
AC AAD33318;  
XX  
DT 01-JUN-2002 (first entry)  
XX  
DE Human C3b/C4b complement receptor like cDNA #1.  
XX  
KW Human; C3b/C4b complement receptor-like molecule; immune system disorder;  
KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;  
KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;  
KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;  
KW transplant rejection; autoimmune disease; ischemic condition; nocturnal;  
KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;  
KW infertility; vasodilator; obesity; cardiac; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 334..9543  
FT /tag= a  
FT /product= "Human C3b/C4b complement receptor like protein  
FT #1"  
FT /transl\_except= (pos:2416..2418, aa:Xaa)  
FT /note= "This translational exception occurs while  
FT decoding the alternative version of human C3b/C4b  
FT complement receptor like protein #1 (AAE20900)"  
XX  
XX W0200210199-A2.

PD 07-FEB-2002.  
 XX  
 XX 24-JUL-2001; 2001WO-US023232.  
 XX  
 XX 02-AUG-2000; 2000US-0222504P.  
 XX 28-NOV-2000; 2000US-00728787.  
 XX  
 PA (AMGE-) AMGEN INC.  
 PI Welcher AA, Elliott GS;  
 XX WPI; 2002-303934/34.  
 DR P-PSDB; AAE20787, AAE20900.  
 XX  
 PT Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic  
 PT acid molecule, useful for treating, preventing and diagnosing rheumatoid  
 PT arthritis, psoriatic arthritis, inflammatory arthritis, and multiple  
 PT sclerosis.  
 PS  
 PS Claim 1; Fig 1; 251pp; English.  
 XX  
 CC The invention relates to a nucleic acid encoding a novel C3b/C4b  
 CC complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like  
 CC polypeptide and nucleic acid molecules may be used to treat, prevent,  
 CC ameliorate, diagnose and/or detect diseases such as immune system  
 CC disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory  
 CC arthritis, osteoarthritis, inflammatory joint disease, autoimmune  
 CC disease, multiple sclerosis, lupus, inflammatory bowel disease,  
 CC transplant rejection, nervous system disorders (e.g. Alzheimer's  
 CC disease), ischemic conditions, metabolic disorders (e.g. obesity and  
 CC diabetes) and infertility. The invention is useful in gene therapy. The  
 CC present sequence is human C3b/C4b complement receptor like cDNA  
 XX  
 XX Sequence 10673 BP; 2690 A; 2711 C; 2574 G; 2691 T; 0 U; 7 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0 Length: 10673  
 Score: 10737.00 Matches: 1864  
 Percent Similarity: 78.37% Conservative: 433  
 Best Local Similarity: 63.60% Mismatches: 567  
 Query Match: 63.21% Indels: 67  
 DB: Gaps: 7  
 US-10-016-248-2 (1-3104) x AAD33318 (1-10673)

QY ArgGluGlySerArgPheHisIsgLYAspThrLeuYsPheGluCySGlnProAlaPhe 310  
 DB CGAGCGGCGAGCACTTTCCTCCATGAGATACACTACCTTGAATCCCGCGCGCTTT 519  
 QY GluLeuValGlyGlnValAlaIleThrCysGlnLysAsnAngIntPSSerAlaLys 330  
 DB GAGCTGTGGGGAGAGATTATCATCCTGTACAGACAATCATGCTGTGGCAACAAG 579  
 QY ProGlyCysValPheSerCysPhePheAsnPheThrSerProSerGlyValValLeuSer 350  
 DB CCCAGCTGTATATTTTCATGTTCTTCACTTCAACGGCAATCATCTGGATTTATCTGCA 639  
 QY ProAsnTYrProGluAspTYrGlyLysAsnIleAsnIleCysValTYrLeuIleLeuAlaArg 370  
 DB CCAATATATCATAGAGAAATATGGAGACAAATACATCTGTCTGTGTTATATCTCGAG 699  
 QY ProGluSerArgIleHisIleuAlaPheAsnAspIleAspValGluProGlnPheAspPhe 390  
 DB CCAAGAACTCGAATTCCTAATCTTAAATGATTTGATGTGAGCTCAATTTGACTTT 759  
 QY LeuValIleLysAspGlyAlaThrAlaGluAlaProValLeuGlyTYrPheSerGlyAsn 410  
 DB CTGCGCGGTCAAGATGATGCAATTTCTGACATTAATCTGCTGGTACTTTTCTGGCAAT 819  
 QY GluLeuProSerSerIleThrSerSerGlyHisIleValAlaArgLeuGluPheGlnThrAsp 430  
 DB GAACTGCTTCCAGCTGCGCAGACAGTGGCAGATTAATCTGCTGGAAATTTCACTGAC 879  
 QY HisSerThrLysArgLysArgLysPheAsnIleThrPheThrThrPheArgHisAsnGluCys 450  
 DB CATTCCACTATCGGCAAGAGGTTCAACATCATCTTACCAACATTTGTGTAATGATGTC 939  
 QY ProAspProGlyValProValAsnGlyLysArgPheGlyAspSerLeuGlnLeuGlySer 470  
 DB CATGATCTCGCAATCTCTTAATAACGACGCTTTGTGACAGCTTTCACTCGGAGC 999  
 QY SerIleSerPheLeuCysAspGluGlyPheLeuGlyThrGlnGlySerGluThrIleThr 490  
 DB TCGCTTCTTCCACTGATGATGATGCTTTGTGTAAGACCAAGGATCCGATCATTAAC 1059  
 QY CysValLeuLysAsnGluGlySerValIleTYrAsnSerAlaValLeuArgCysGluAlaPro 510  
 DB TGATATCTGCAAGCGGGAAGTGTGAGCTCCACCGTCCCGCTGTGAAGCTCCA 1119  
 QY CysGlyGlyHisIleuThrSerProSerGlyThrIleLeuSerProGlyTYrProGlyPhe 530  
 DB TGTGTGTGACATCTGACAGCGTCCAGCGAGTCAATTTTGCTCTGTGATGCCAGGATAT 1179  
 QY TyrIleAspAlaLeuSerCysAlaIleTYrValIleGluAlaGlnProGlyTYrProIleLys 550  
 DB TATPAGGATTTCTTACATTTGTGAATGATATTAAGCAACCAACGAGCCACTCTATCAAA 1239  
 QY IleThrPheAspArgPheLysThrGluValAlaAsnTYrAspThrLeuGluValAlaArgAspGly 570  
 DB ATAACCTTTACAGATTTCAGACAGAGTCAATTAATGACACTTGGAGGTCAAGATGGG 1299  
 QY ArgThrTYrSerAlaProLeuIleGlyValTYrHisGlyThrGlnValProGlnPheLeu 590  
 DB CCAAGCAAGTTCGTCCTCCAGATCGGCACTACCAAGCCAGCAAGCCAGTTCCTC 1359  
 QY IleSerThrSerAsnTYrLeuTYrLeuLeuPheSerThrAspLysSerHisSerAspIle 610  
 DB ATCAGACCGGGAATTCATGATCTGCTATTCACCTACCAACACCGCTCCAGATTC 1419  
 QY GlyPheGlnLeuArgTYrGluThrIleThrLeuGlnSerAspHisCysLeuAspProGly 630  
 DB GGCTTCTCATCCACTATGAGATGAGTGAACGCTTGAGTCCGATTCCTGCAACCGCGG 1479  
 QY IleProValAsnGlyLysArgHisIsgLYAsnAspPheTYrValGlyAlaLeuValThrPhe 650  
 DB ATCCCTGTGAACGCAATCCCAAGCGGTGAGACTTTGGCATCGATCCCAAGTGAATTC 1539  
 QY SerCysAspSerGlyTYrThrLeuSerAspGlyGluProLeuGluCysGluProAsnPhe 670

```
Db 1540 AGCTGTGACCCGGGGTACACATAAGTACACAGAGCCCTCGTGTGAGAGAACAC 1599
Qy 671 GINTPSErAglAlaLeuProSeCygluAlaLeuCyglYglYpHeIIeGInglYser 690
Db 1600 CAGTGGAAACACGCGCTTGCCAGCGTCCAGCTCTATGTGAGAGGCTCATCTCAAGGAG 1659
Qy 691 SerGlyThrIleLeuSerProGlyPheProAspPheTyrrProAsnAsnLeuAsnCysThr 710
Db 1660 AGTGAACAGTCTCTTCTCTCGGTTTCCAGATTTTATCCAACTCTCTAAACGCGACG 1719
Qy 711 TrpIleIleGInThrSerHieGlyValpHePheThrPheHISThrPheHISLeu 730
Db 1720 TGGACCATTTGAAGTGTCTCATGGGAAGAGATTCAAAATGATCTTCAACCTTCACTT 1779
Qy 731 GluSerGlyHISAspTyrrLeuLeuIleThrGluAsnGlySerPheThrGlnProLeuArg 750
Db 1780 GAGAGTTCACACGACTATTTATCTATCAACAGAGATGGAAGTTTTCGAGCCGTTGCC 1839
Qy 751 GlnLeuThrGlySerArgLeuProAlaProIleSerAlaGlyLeuTyrrGlyAsnPheThr 770
Db 1840 AGGCTCACCGGGGTGGGTGTTCCCTCATACATCAAGGACAGGCTGTTGAAACTTCACT 1899
Qy 771 AlaGlnValArgPheIleSerAspPheSerSerTyrrGlyPheAsnIleThrPhe 790
Db 1900 GCCACGCTTCGGTTTATATACAGACTTCTCAATTCGTACAGAGGCTTCAATATCACTTT 1959
Qy 791 SerGlyTyrrAspLeuGluProCygluGluProGluValProAlaTyrrSerIleArgLys 810
Db 1960 TCAGAAATATACCTCGAGCCATGATGATCTCGAGATCTCTCGCTTCACCCCAAGAAATT 2019
Qy 811 GlyLeuGlnPheGlyValGlyAspThrLeuThrPheSerCySpheProGlyTyrrArgLeu 830
Db 2020 GGTTCATCTTGGTGTGGAGACTCTCTGACGTTTTCGCTTCCTGGGATATCTTTA 2079
Qy 831 GluGlyThrAlaArgIleThrCygluGlyValArgArgLeuThrSerSerProLeu 850
Db 2080 GAAGGTGCCRCAAGCTTACCTCGCTGGGCGGGGCGCGCTGTGTGAGTGCACCTCTCG 2139
Qy 851 ProArgCyValAlaGlyCyglYAsnSerValThrGlyThrGlnGlyThrLeuLeuSer 870
Db 2140 CCAAGGTGTGTGCGCATGTGAGCAAGTGTCAAGAAAGAAATGAAAGAAATTAATGCTGT 2199
Qy 871 ProAsnPheProValAsnTyrrAsnAsnAsnHISGlyCysIleTyrrSerIleGlnThrGln 890
Db 2200 CCAAAATTTTCCATCCATATGATTAATTAACATGAGTATCTATAAATAGAAACAGA 2259
Qy 891 ProGlyValGlyIleGlnLeuLysAlaArgAlaPheGlnLeuSerGlyLysAspValLeu 910
Db 2260 GCCGGCAAGGGCATCCACTTACAGAACAGAGCTTCAGCTGTGTAAGAGAAATACTCTA 2319
Qy 911 LysValTyrrAspGlyAsnAsnAsnSerAlaArgLeuLeuGlyValPheSerHISserGln 930
Db 2320 AAGGTATATGATGAAAGAACAGTCTCTACGTCACATGGGACGTTCACTAAATAATGA 2379
Qy 931 MetMetGlyValThrLeuAsnSerThrSerSerSerLeuThrLeuAspPheIleThrAsp 950
Db 2380 CTTCTGGGGCTGATCTCTAAACAGACATCCATCACTGTGGCTAGAGTTCAACACCAAT 2439
Qy 951 AlaGlnAsnThrSerIleGlyPheGlnLeuLysAlaPheSerSerPheGlnLeuLeuLysCys 970
Db 2440 GGAATCGACACCAACCAAGTTTCACTCACTATACAGATTTGATGTGTAATAATGT 2499
Qy 971 GluAspProGlyThrProLysPheGlyTyrrLysValHISAspGluGlyHISpHeAlaGly 990
Db 2500 GAGGATCCGGGCTATCCCTAACTACGCTATAGGATCCGATGAGGCCACTTTACCGAC 2559
Qy 991 SerSerValSerPheSerCySpAspProGlyTyrrSerLeuArgGlySerGlnGluLeu 1010
Db 2560 ACTGTAATGTTCTGTACAGTTCACCCGGGGTACGGCATGTGAGGACGAACACCTCGACC 2619
Qy 1011 CysLeuSerGlyGluArgArgThrTrpAspArgProLeuProThrCyValAlaGlnCys 1030
Db 2620 TGTTTAGTGGAGACAGAGAGTGTGGACAAACCACTACTTGTGATAGCGGAATGT 2679
Qy 1031 GlyGlyThrValArgGlyGluValSerGlyGlnValLeuSerProGlyTyrrProAlaPro 1050
Db 2680 GGTGTCAATCATCAGTACAGCCACATCAGACGAATTTGTCCCTGTGCTATCCAGTCCG 2739
Qy 1051 TyrrGlnHISAsnLeuAsnCysIleThrThrIleGluAlaGluAlaGlyCysThrIleGly 1070
Db 2740 TATGACAAACCTCCACTGACCTGGATTAATAGAGGACAGCCAGAAAGACATTAAC 2799
Qy 1071 LeuHISpHeLeuValPheAspThrGlnGluValHISAspValLeuAspGlyLeuThrAspGly 1090
Db 2800 CTTCATTTATTTGTTTTCACACGAGATGCTCACAGATCTCTAAGGTCTTGGAGCGGG 2859
Qy 1091 ProValGlnSerGlyValLeuLeuLysGlnLeuSerGlyProAlaLeuProLysAspLeu 1110
Db 2860 CCGGTGACAGATGACATCTGCTGAAGAGTGAAGTGGCTCGGCCCTTCGGAGGACATC 2919
Qy 1111 HISerThrPheAsnSerValValLeuGlnPheSerThrAspPhePheThrSerLysGln 1130
Db 2920 CACAGACCTTCACTCACTCACTGACATTCAGACGACTTCTTATCATCGAAGTCT 2979
Qy 1131 GlyPheAlaIleGlnPheSerValSerThrAlaThrSerCySpAsnAspProGlyIlePro 1150
Db 2980 GGCTTCTCATCATCTTCTCCACTCATTTGACGCCACTGTAAAGATTCAGATATGCC 3039
Qy 1151 GlnAsnGlySerArgSerGlyAspSerTrpGluAlaGlyAspSerThrValPheGlnCys 1170
Db 3040 CAAATATGACCCGCTATAGAGACAGAGAGGCTGAGACACCGTCACATTCAGTGT 3099
Qy 1171 AspProGlyTyrrAlaLeuGlnGlySerAlaGlnIleSerCyValLysIleGluAsnArg 1190
Db 3100 GACCTGTGCTACCTCCAAAGACACCAAAATCACTGTGTGACATAATTAATTCGG 3159
Qy 1191 PhePheTrpGlnProSerProProThrCysAlleAlaProCyglYAspLeuThrGly 1210
Db 3160 TTTCTTTGGCAACGACACCTCTCATAGATGATGCTGTTGGAGGGAATCTAGCGGC 3219
Qy 1211 ProSerGlyValIleLeuSerProAsnTyrrProGluProTyrrProProGlyLysGlnCys 1230
Db 3220 CCAGAGGTGTATTTGTGTACCCAACTACCAACGCGGTATCTCTCGGGAAGAAATGT 3279
Qy 1231 AspTrpLysValThrValSerProAspTyrrValIleAlaLeuValPheAsnIlePheAsn 1250
Db 3280 GACTGGAAGTAAGTAAGTAACCCGAGCTTTGTCTGATGATTAATCAAAAGTTTCAAC 3339
Qy 1251 LeuGlnProGlyTyrrAspPheLeuHISIleTyrrAspGlyArgAspSerLeuSerProLeu 1270
Db 3340 ATGAGGCCACCTATGACTTCTCTACATCTATGAAAGGGAATTCACAGCCCCCTC 3399
Qy 1271 IleGlySerPheTyrrGlySerGlnLeuProGlyArgIleGlySerSerSerAsnSerLeu 1290
Db 3400 ATTGGAGATTACAGAGGCTCTCAGAGCCCAAGAAAGTAAGATACCGGAACAGCTCG 3459
Qy 1291 PheLeuAlaPheArgSerAspAlaSerValSerAsnAlaGlyPheValIleAspTyrrThr 1310
Db 3460 TTTCTGGGATTTGGAGATGATGCTCGGTGGCCCTTCAGAGTTCCGCATTTAAATTAA 3519
Qy 1311 GluAsnProArgGlnSerCySpheAspProGlySerIleLysAsnGlyThrArgValGly 1330
Db 3520 GAGAAACACGAGGAGGTTGTTTGAACCCAGGAATTAATGATGAGGAACAAAGATTGGA 3579
Qy 1331 SerAspLeuLysLeuGlySerSerValThrTyrrCysHISGlyGlyTyrrGlnValGlu 1350
Db 3580 ACAATCTTCAAGCTTGCTCACAATACATCACTGAGTGAATCTGTGCTTAAGATTTCTT 3639
Qy 1351 GlyThrSerThrLeuSerCysIleLeuGlyProAspGlyLysAspProValTrpAsnAsnPro 1370
Db 3640 GACCCCTCATCATCACTGCTGTGATTTGGGCTGATGGGAACCTTCCTGGACCAAGTG 3699
Qy 1371 ArgProValCysThrAlaProCyglYglnGlyValGlySerAspGlyValLeu 1390
Db 3700 CTGCTCTCTGCAATGCTCTGTTGAGGCGCAGTACACGGGATCAGAAAGGATGATTTA 3759
```



```
Db 5920 AACATTACCATCTTTGGACACATTTCAAGTGAAGAGCTTTGATGACCTGAAAGTG 5979
Qy 2131 PheAspGlyProSerGlyGlnSerProLeuLeuYalaLeuSerGlyYAsnTyrSerAla 2150
Db 5980 TTTGATGGTTCTTCTGGGCAAAAGTCTCTAGTACGTCTTAAGTGGAAATCATATCAAG 6039
Qy 2151 ProLeuLeuValThrSerSerAsnSerValTyrLeuArgTyrPsrSerAspHisAla 2170
Db 6040 CAATCAAAATTTTACAGCAGAGATATCATGTTATATCTCGCTGGTCTCAGTACAGCAAGCC 6099
Qy 2171 TyrAsnArgGlySerGlyPheYalaIleArgTyrSerAlaProTyrCySerLeuProArgAla 2190
Db 6100 ACCAGTAAGAAAGATTCAGATTCGCTATGACACACTTACCTGACATTTGACCCACC 6159
Qy 2191 ProLeuHisGlyPheIleLeuGlyGlnThrSerThrGlnProGlyGlySerIleHisPhe 2210
Db 6160 CTGAAGATGGGGGTATTTCAACAGACTGACAGACGGGTGGACCAAGTGCATATAT 6219
Qy 2211 GlyCyAsnAlaGlyTyrArgLeuValGlyYHisSerMetAlaIleCyThrArgHisPro 2230
Db 6220 TTTTTCAGACCTGGATATCCGATATGTCGGCCACAGCAATGCACTTATACAAACCA 6279
Qy 2231 GlnGlyTyrHisLeuTyrPsrSerGlnAlaIleProLeuCyGlnAlaLeuSerCyGlyLeu 2250
Db 6280 CTGGCATGTACACAGTGGGACTCCCTCAGCCACTGCGACGGTGTCTGTCGTGGATC 6339
Qy 2251 ProGlnAlaProLyAsnGlyMetValPheGlyYLeuGlyTyrThrValGlyThrIleVal 2270
Db 6340 CCAAGATCCCGAGAAACGGTTCTTATCCGGGAACAGTTCACTTGGACAGTAAAGTG 6399
Qy 2271 ValTyrSerCySerGlyGlyTyrHisLeuGlnAlaGlyYalaGlnAlaThrAlaGlyCy 2290
Db 6400 GTCTATGAATGTCATGAGGCTTCAGCTTCAAGCTTGAATCCAGCAGCAAGAACGCCGTGT 6459
Qy 2291 LeuAspThrGlyLeuTyrPsrSerAsnValProProGlnCyValProValThrCy 2310
Db 6460 CAAAGAAATGGCTGTGGATTAACAAGGGAGCCGCCACGTGTAAAGCCGCTCGCTTC 6519
Qy 2311 ProAspValSerSerIleSerValGlnHisGlyYArgTyrPsrLeuIlePheGlnThrGln 2330
Db 6520 CCAAGCATTAAGCTCAGCTCTCAGAACATGTCATGAGGCTGTCTCAGAGATCTTGG 6579
Qy 2331 TyrGlnPheGlnAlaGlnLeuMetLeuIleCyAspProGlyTyrTyrTyrThrGlyGln 2350
Db 6580 AATGATGACGCTCTCAAGATATGCTGAGCTGCAGTCTGTACTTACTTAAGAGCTGG 6639
Qy 2351 ArgValIleLeuArgGlyGlnAlaGlnGlyYLeuTyrPsrLeuGlyYAspSerThrProThrCy 2370
Db 6640 AGGCTCTGGGTCAGGCAAGGCAATGGAAGCATAGAGATAGAGGCAAGGCTCT 6699
Qy 2371 ArgIleIleSerCyGlyGlnLeuProIleProProAsnGlyYHisArgIleGlyThrLeu 2390
Db 6700 CGAGTTATCTCGTGGAGAGCTTTCTTCCCCCAAGGCAACAGATTGGAAGCTTG 6759
Qy 2391 SerValTyrGlyYalaThrAlaIlePheSerCyAsnSerGlyTyrThrLeuValGlySer 2410
Db 6760 ACGATTATAGGGCCACAGCTATATTTAGCTGCAACCGGCTACACGCTTGGGGTCT 6819
Qy 2411 ArgValArgGlyCyMetAlaGlnGlyYLeuTyrPsrGlySerGlnValArgCySerLeuAla 2430
Db 6820 CATGTAGAGAGCTTTGGCAATGGGCTCTGGAGGGGCAAGCACTGAGTGTGGCT 6879
Qy 2431 GlyHisCyGlyThrProGlnProIleValAsnGlyHisIleAsnGlyGlnLeuTyrSer 2450
Db 6880 GGCACATGCGGTTCCCAAGCCGATTTGAAACGGTCATATAGTGAAGATGGCTTCACT 6939
Qy 2451 TyrArgGlySerValValTyrGlnCyAsnAlaGlyPheArgLeuIleGlyMetSerVal 2470
Db 6940 TACAGAGACAGGTGGTTTACAGAGCAATCGTGTTCGGCTTGGGAACTTCCGCTG 6999
Qy 2471 ArgIleCyGlnGlnAspHisIleTyrPsrGlyYLeuThrProPheCyValProIleThr 2490
Db 7000 AGGATATGCCCTGCAGACCAAGAGTGTGACAAACCCCTGTGTCTCCCATCACA 7059
Qy 2491 CyGlyYHisProGlyYAsnProValAsnGlyYLeuThrGlnGlyYAsnGlnPheAsnLeuAsn 2510
Db 7060 TGTGTACACCTCGAAGAACCTCGCCACGGAATTCACATTAAGGACAGTCAACTCAAT 7119
Qy 2511 AspValIleValPheValCyAsnProGlyTyrMetAlaGlnGlyYalaIleArgSerGln 2530
Db 7120 GATGTGTAAATTTTCACTGCAACCGGCTATTTGTGCGACAGGGCGTGTGAGCCAG 7179
Qy 2531 CyLeuAlaSerGlyGlnTyrPsrSerAspMetLeuProThrCyAsnGlyIleLeuAsnCyThr 2550
Db 7180 TGTGGAGCAAGCGGCAAGTGAAGTACCCTGTGCCACGCTGTGTGTGATGACTGTCT 7239
Qy 2551 AspProGlyYHisGlnGlnAsnSerValArgGlnValHisAlaSerGlyProHisArgPhe 2570
Db 7240 GATCAGGCTTGTGGAATAATGCATGCTGTCAGGCAACAGAACTTCCCTGAGGTTT 7299
Qy 2571 SerPheGlyThrThrValSerTyrArgCyAsnHisGlyPheTyrLeuLeuGlyThrPro 2590
Db 7300 GAGTATGAAATGAGATCTGTACCATTCAGAAAGGAATTTTACTTCTGTGGATCTTCA 7359
Qy 2591 ValLeuSerCyGlnGlyYAspGlyYThrTyrAspArgProArgProGlnCySerLeuVal 2610
Db 7360 GCTTGACCTGTATGGCAAAATGCTTATGGACCAATCCCTGCGCAAGTGTGGCTTAT 7419
Qy 2611 SerCyGlyYHisProGlyYSerProProHisSerGlnMetSerGlyYAspSerTyrThrVal 2630
Db 7420 TCGTGTGACACCCAGGGGTCTCTGCAACCCGCTCTCAGTGGAGACTGTATCTAT 7479
Qy 2631 GlyAlaValIleArgTyrSerCySerIleGlyYAsnArgThrLeuValGlyYAsnSerThrArg 2650
Db 7480 GCGCCGCTGTGCACTACTCTGCAAGAGGACGAGAACCTCTATAGGCAACGACACGAGA 7539
Qy 2651 MetCyGlyLeuAspGlyYHisTyrThrGlyYSerLeuProHisCySerGlyYThrSerVal 2670
Db 7540 GTGTCCAGAAAGACAGTCACTGAGCGGGCACTGCCCACTGCACAGAAATATCT 7599
Qy 2671 GlyValCyGlyYAspProGlyYIleProAlaHisGlyYIleArgLeuGlyYAspSerPheAsp 2690
Db 7600 GGAATTCGTGTGATCCCGGAGACCCAGCAATGGTCTCGGCTTGTGATGACTTTAAG 7659
Qy 2691 ProGlyThrValMetArgPheSerCyGlnAlaGlyYHisValLeuArgGlyYSerSerGly 2710
Db 7660 ACAAAGATCTTCTCCGCTTCTCTGTGAATGGGGACCAAGCTGAGGGCTCTCCCTGAA 7719
Qy 2711 ArgThrCyGlnAlaAsnGlyYSerTyrPsrGlyYSerGlnProGlyCyGlyYValIleSer 2730
Db 7720 CGCAGTGTCTTGTCAATGGGTCAATGATGATGAGCACTGCAGCGCGGTGTGAGCGGTCT 7779
Qy 2731 CyGlyYAsnProGlyYThrProSerAsnAlaArgValValPheSerAspGlyYLeuValPhe 2750
Db 7780 TGTGGCAACCTGGACACCCACCAAGGAATGATTTCTAGTATGATGATGATGATGATGAT 7839
Qy 2751 SerSerSerIleValTyrGlyCyArgGlyGlyYThrTyrYalaThrGlyYLeuLeuSerArg 2770
Db 7840 TCCAGCTCGGCATCTATGCTGCTGTGGAGGCTCAACAAGACCTGAGGCTCATACACGG 7899
Qy 2771 HisCySerValAsnGlyYThrTyrThrGlyYSerAspProGlyCySerLeuValIleAsnCy 2790
Db 7900 CATTGCACAGCCAATGGGACTGGACAGGCACTGCTCCGACATGCAAAATTAATGATGT 7959
Qy 2791 GlyAspProGlyYIleProAlaAsnGlyYLeuArgLeuGlyYAsnAspPheArgTyrAsnIle 2810
Db 7960 GGGGATCCAGGACACATAGCAAAATGGATTCAGATTGGGACCACTTCAACAAAG 8019
Qy 2811 ThrValThrTyrGlnCyValProGlyYThrMetMetGlnSerHisArgValSerValLeu 2830
Db 8020 ACTGTGAGTATAGTGTAACTCAGGCTATGTATGGAACAGATCATCCGCACTAT 8079
Qy 2831 SerCyThrTyrAspArgThrTyrPsrGlyYThrYAspProValCyYalaLeuMetCy 2850
Db 8080 CGCTGTACAAAGACGCGCAGGTGAATCCGAGCAAACTGTCTGCAAAACCGCTGTGTGT 8139
```



QY 307 GlnProAlaPheGlnLeuValGlyGlnLysAlaIleThrCysGlnLysAsnGlnIleTrp 326  
 DB 253 CAGCGAGCTTTGAGCTGTAGAGAGAGATGATTTAGCTGCCAAGAAACAACAGAGG 312  
 QY 327 SerAlaLysLeuProGlyCysValPheSerCysPhePheAsnPheThrSerProSerGly 346  
 DB 313 TCCGGCAACAGCAAGCTGTGTGTTTCACTGTTTCACTTCACTGAGCGCTCTCGG 372  
 QY 347 ValValLeuSerProAsnThrProGlnAspTyrGlyAsnHisLeuHisCysValIleTrp 366  
 DB 373 ATCATCTGTCTCCCAACTATCTGAGAGAAATGAGCAACATGAAATGTGTGTGTG 432  
 QY 367 IleLeuAlaArgProGlnSerArgIleHisLeuAlaPheAsnAspIleAspValGlyPro 386  
 DB 433 ATTATATCTGAGCCCGGAGCGGAGTCACTCATCTTCAATGATTTGAGTGTGAGCT 492  
 QY 387 GlnPheAspPheLeuValIleLysAspGlyAlaThrAlaGlnAlaProValLeuGlyThr 406  
 DB 493 CAGTTGACTTCTTGGCTGCGATCAAGATGATGAGATTTCGACATCAACAGTCTCGGACT 552  
 QY 407 PheSerGlyAsnGlnLeuProSerSerIleThrSerSerGlyHisValAlaArgLeuGly 426  
 DB 553 TTCTGTGCAATGAGTGTCTGACAGCTGCGCAGCTGAGCAATAGTACGCTGTGAG 612  
 QY 427 PheGlnThrAspHisSerThrGlyLysArgGlyPheAsnIleThrPheThrThrPheArg 446  
 DB 613 TTTCAGTCCGATCACTTACCAACGAGGAGGTTCAACATCATATACCAACCACTTTGCT 672  
 QY 447 HisAsnGlnCysProAspProGlyValProValAsnGlyLysArgPheGlyAspSerLeu 466  
 DB 673 CAGAACAGTGTATGACCTTGGAGATCCCTGTGATGAGACGCGCTTTGAGACAGGTTT 732  
 QY 467 GlnLeuGlySerSerIleSerPheLeuCysAspGlnGlyPheLeuGlyThrGlnGlySer 486  
 DB 733 CTGCTGGAGAGTTCTGTGCTCTTCCACGTGATGATGCTTTGAGAGATCAGGTTCT 792  
 QY 487 GlnThrIleThrCysValLeuLysGlnGlySerValIleTrpAsnSerAlaValLeuArg 506  
 DB 793 GAGCTATCATCATCATCTTGCAGAGATGAAACCTGTCTGAGACTCTACTGCTCCCTCG 852  
 QY 507 CysGlnAlaProCysGlyGlyHisLeuThrSerProSerGlyThrIleLeuSerProGly 526  
 DB 853 TGTAACTCTCTTGTGTGGGATCTGACAGCTTTTATGTGGGTATATTAACCTTCAGGA 912  
 QY 527 TrpProGlyPheTyrLysAspAlaLeuSerCysAlaIleTrpValIleGlnLysGlnVal 546  
 DB 913 TGGCAGAGATATTAACAAGATCTTTAATTGGAAATGGGTCACTGAGCAACAGGA 972  
 QY 547 TyrProIleLysIleThrPheAspArgPheLysThrGlnValAsnTyrAspThrLeuGln 566  
 DB 973 CATTCATCAAAATATACATTTGACAGGTTCACAGCAAGATCAATATATATCTCGGAA 1032  
 QY 567 ValArgAspGlyArgThrTyrSerAlaProLeuIleGlyValTyrHisGlyThrGlnVal 586  
 DB 1033 GTCGGGATGGGCAACAGCTCATCCCACTGATGGGAGATGACATGACCCAGGCT 1092  
 QY 587 ProGlnPheLeuLeuSerThrSerAsnTyrLeuTyrLeuLeuPheSerThrAspLysSer 606  
 DB 1093 CCAAGATCTCTATCAAGCAACAGGAACTACATGTACTCTGTTTACCACTGACAGCAG 1152  
 QY 607 HisSerAspIleGlyPheGlnLeuArgTyrGlnIleThrLeuGlnSerAspHisCys 626  
 DB 1153 CCGGCAAGTGTGTCTCTCATCCATGAGAGGTGATCTTGAATCTGATCTCTGCT 1212  
 QY 627 LeuAspProGlyLysProValAsnGlnLysGlnArgHisGlyAsnAspPheTyrValGlyAla 646  
 DB 1213 CTGGACCCGCGGATCTCTTAATGTGTCATCGGCAATGAGCAATCTTGTGATCATGACT 1272  
 QY 647 LeuValThrPheSerCysAspSerGlyTyrThrLeuSerAspGlyGlnProLeuGlnCys 666  
 DB 1273 AAGATGACCTTCACTGATGACCTCGGTAACGCTCATGATGAGATCCCTCATCTGT 1332

---

QY 667 GlnProAsnPheGlnIleTrpSerArgAlaLeuProSerCysGlnAlaLeuCysGlyGlyPhe 686  
 DB 1333 GAGAGAACCATCATGAGGAACAACGCTTGCCAGCTGTGATGAGCTGTGTGAGGCTAC 1392  
 QY 687 IleGlnGlySerSerGlyThrIleLeuSerProGlyPheProAspPheTyrProAsnAsn 706  
 DB 1393 ATCATGAAAGATGAGCTGTCTTTTACAGAGATTTCCAGACTTTTATATCAACTCT 1452  
 QY 707 LeuAsnCysThrTrpIleIleGlnThrSerHisGlyLysGlyValPhePheThrPheHis 726  
 DB 1453 CTGAATCTGATACATGAGCATTTGATGATCTCTCATGCGAAGGATGACATGAAATTTCC 1512  
 QY 727 ThrPheHisLeuGlnSerGlyHisAspTyrLeuLeuIleThrGlnAsnGlySerPheThr 746  
 DB 1513 ACCTTTCACTTGAAGATTCACAGCATATTTGTGATACAGAGATGGAGTTTCTCA 1572  
 QY 747 GlnProLeuArgLysLeuThrGlySerArgLeuProAlaProIleSerAlaGlyLeuTyr 766  
 DB 1573 GAGCGCGTACAGGCTCATGTGGTCTGCTGCTCACACCATTAAGGCTGGCTGTGTT 1632  
 QY 767 GlyAsnPheThrAlaGlnValArgPheIleSerAspPheSerMetSerTyrGlnGlyPhe 786  
 DB 1633 GGAAACTTCACTGCGCAACTAGAGTTTCACTCTGACTTCTCATCTCTATGAGGCTTC 1692  
 QY 787 AsnIleThrPheSerGlyThrAspLeuLysProCysGlnGlnProGlnValProAlaTyr 806  
 DB 1693 AACATTAGCTTTGAGAAATATGACCTTGAACCTGTGTGACCCCTGAGATCCCTGCTTAC 1752  
 QY 807 SerIleArgLysGlyLeuGlnPheGlyValGlyAspThrLeuThrPheSerCysPhePro 826  
 DB 1753 AGTCCAGAAATGTGGTTCAGTGGGTGTGGGTGACACCTTGCTTCACTGCTGCTCAG 1812  
 QY 827 GlyTyrArgLeuGlnGlyThrAlaArgIleThrCysLeuGlyGlyArgArgLeuTrp 846  
 DB 1813 GGATACCGCTTATGAGAGTGCAACCAAGCTTACCTGCTGTGGGAGCGCGAGTGTGG 1872  
 QY 847 SerSerProLeuProArgCysValAlaGlnCysGlyAsnSerValIleThrGlyThrGlnGly 866  
 DB 1873 AGTGACCTCTGCGCAAGTGTGTGTGTGATGTGAGCAACGCTCAAGAAATGAAAGGA 1932  
 QY 867 ThrLeuLeuSerProAsnPheProValAsnTyrAsnAsnAsnHisGlyCysIleTyrSer 886  
 DB 1933 ACATTACTCTCTCCAAATTTCCATCCATATATGATATATACATGATGATATATAA 1992  
 QY 887 IleGlnThrGlnProGlyLysGlyIleGlnLeuLysValArgAlaPheGlnLeuSerGln 906  
 DB 1993 ATGAAACAGAGCGGAAAGGGGATCATCTCAGAGCCGGAACCTTCCAACTTTCGAA 2052  
 QY 907 GlyAspValLeuLysValTyrAspGlyAsnAsnAsnSerAlaArgLeuLeuGlyValPhe 926  
 DB 2053 GGAGACACTTAAAGGTTTATGATGAGAAAGACAGCTCTCGAGGTCACTGGAGTCTTC 2112  
 QY 927 SerHisSerGlyMetMetGlyValThrLeuAsnSerThrSerSerIleTrpLeuAsp 946  
 DB 2113 ACAAGAAATGAACTGAGGGGCTGTGCTTAAACAGACCTTCCAAACCTGAGGCTGAG 2172  
 QY 947 PheIleThrAspAlaGlnAsnThrSerLysGlyPheGlnLeuHisPheSerSerPheGln 966  
 DB 2173 TTCAACTCTAACGGGTACATATCCGCCCAAGCTTTCAGCTCACTCAACCAATTTTGAC 2232  
 QY 967 LeuIleLysCysGlnAspProGlyThrProLysPheGlyTyrLysValHisAspGlnGly 986  
 DB 2233 CTAGTAAATGATGAGATCCAGGCACTCCCTAATAGGTGATACAGATCCGAGATGATGT 2292  
 QY 987 HisPheAlaGlySerSerValSerPheSerCysAspProGlyTyrSerLeuArgLysSer 1006  
 DB 2293 CACTTACAGACACTGTGTCTTCAACACTGCAACCCAGGCTACCAATGATGATGACAC 2352  
 QY 1007 GlnGlnLeuLeuCysLeuSerGlyGlnArgArgThrTrpAspArgProLeuProThrHis 1026  
 DB 2353 AGTACCTGACCTGCTGATGTGGGAGCCGAAAGGTGTGGAAACAAACTATGCTTCTGT 2412  
 QY 1027 ValAlaGlnCysGlyGlyThrValArgGlyGlnValSerGlyGlnValLeuSerProGly 1046

|    |      |   |      |
|----|------|---|------|
| Db | 2413 | GTGGCGGAATATGGTGGTCTCTGTCATGACACCATGAGACCGATATCTCTCTGCG         | 2472 |
| Qy | 1047 | TYRPROALAPROYTGLNHIAASNLDSANCYSLIETPRTHILEGLUALAGLUNLAGLY       | 1066 |
| Db | 2473 | TACCTGGCCCATATGACAAACACCTTATGTGACATTGGACCATAGAGCGTGCATCTGGC     | 2532 |
| Qy | 1067 | CysrthriIegLYLeuHIApHeuValIpheasrThrgiUGluValHIAspValIleuArg    | 1086 |
| Db | 2533 | AAGACCAACAGCCCTCCATTTTCAATGTGTGTGAACCTGAAGAGCGGCACGACATCTCCAG   | 2592 |
| Qy | 1087 | ILetRPasBpLYProValIGluSerGLYValIleuLeuLYSGluIleuSerGLYProAlaIeu | 1106 |
| Db | 2593 | GTCTGGAGTGGTTCAGTGACGACGACAACTCCGTGAAAGAGGTGAGCGGCTCGCCCTT      | 2652 |
| Qy | 1107 | ProLYAspLeuHIAseRThrPheAsnSerValIleuGlnPheSerThrAspPhe          | 1126 |
| Db | 2653 | CCTGAGGACATCCACAGCACCTTCAACTCGCTCAACCCCTGCAGTTGATAGGACTTCTTC    | 2712 |
| Qy | 1127 | ThrsrLYSGlnGLYpHeaIILEgInPheSerValSerThraIarHsrCYasnaSp         | 1146 |
| Db | 2713 | ATCGCAAGTCCGGCTCTCCATCCAGATCTCTCACTTCCATATGCACTCACCTGCATATAC    | 2772 |
| Qy | 1147 | ProGLYLeuProGlnaNGLYSerThrgSerGLYAspSerTPGLUNLAGLYAspSerThr     | 1166 |
| Db | 2773 | CTTGAGATGCTCGAGAAATGAAACCCGCTATAGGTGACACCGGGAACCTGGAGACACATC    | 2832 |
| Qy | 1167 | ValPheGLINCysAspPProGLYTYrValaLeuGlnGLYSerAlaGLYLeuSerCYeValLYs | 1186 |
| Db | 2833 | ACCTTCCAGTGTGACCTCGATACAGCTCCAAAGGCAAGCAAGTACTGTGTGTGAG         | 2892 |
| Qy | 1187 | ILEgluAnaArgPhePheTrpGlnProSerProProThrCYsIleAlaProCYsGLYGLY    | 1206 |
| Db | 2893 | CTTAAACAACCCCTCTCTCTGGCAACCAAGCCCTCGCATATAGCTGTGTGTGTGG         | 2952 |
| Qy | 1207 | AspLeuThrgLYProSerGLYValIleuLeuSerProAsnTYrProGlnProTYrProPro   | 1226 |
| Db | 2953 | AATCTGACAGGCCCCGTGAGGTATTTATTCCTCCAACTACCAACACCAACCACTTCTCT     | 3012 |
| Qy | 1227 | GLYLYSGIUCYAspPTPLpLYValThrValSerProAspTYrValIleAlaLeuValPhe    | 1246 |
| Db | 3013 | GGGAAGAGTGTGACTGAGAAATTAAGGTGAACCACTTGTCTATTCCTTTAATATTC        | 3072 |
| Qy | 1247 | AsnIlePheAsnLeuGlnProGLYTYrAspPheLeuHIAseLYAspGLYArgAspSer      | 1266 |
| Db | 3073 | AAAGATTTACATGAGAGCCAAAGTTACGACTTCCGTACATCTTAATGAAGGAGACTCC      | 3132 |
| Qy | 1267 | LeuSerProLeuIILEGLYSerPheTYrGLYSerGlnLeuProGLYArgILEGLYSerSer   | 1286 |
| Db | 3133 | AACAGCCACATGATCGGAAGCTTCCAGGGTCTCAAGCCCAAGAGGATTTGAAGACGCT      | 3192 |
| Qy | 1287 | SerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerValSerAsnAlaGLYPheVal    | 1306 |
| Db | 3193 | GGTAACAGCCCTCTCTCTGCACTTCAAGAGTATGCTCTGTGGCCCTCTCCGGATTGCC      | 3252 |
| Qy | 1307 | ILeasPTYrThrgIuAsnProArgLYSerCYsPheAspProGLYSerILELYAsnGLY      | 1326 |
| Db | 3253 | ATTGAATTTAAAGAGAAACCAACGGGAGACTTGCTTTGACCTCGGGAACATATGACGGC     | 3312 |
| Qy | 1327 | ThrTrgValGLYSerAspLeuLYLeuGLYSerSerValIThrTYrCYsHIAGLYGLY       | 1346 |
| Db | 3313 | ACAAGGATTTGAAACGGACTTTTAAGCTGGGCTCTCAAGTTACTCATATGACTGACTGGT    | 3372 |
| Qy | 1347 | TYrGLUValIGluGLYThsrThrLeuSerCYsAlIleuGLYProAspGLYLYsProVal     | 1366 |
| Db | 3373 | THCAAGATTTGGATCCCTCATCTCAATTAAGTGTGTGACAGAGGCGCTGATGGAAAGCTTCC  | 3432 |
| Qy | 1367 | TrpAsnAsnProArgProValCYsETHrAlaProCYsGLYGLYGLYTYrValIGLYSerAsp  | 1386 |
| Db | 3433 | TGGGACCGCGGACATGCGCTGCTGCCAAGCAACCTGTGGAGCCAAATACAGGCTCGGAG     | 3492 |
| Qy | 1387 | GLYValValLeuSerProAsnTYrProGlnAsnTYrThsrGLYGLYIleCYeLeuTYr      | 1406 |

|   |      |   |      |
|---|------|---|------|
| D | 3493 | GGGGTAGTTTGTGCACCAAACTACCTCCCTAATACATCAACGGCTGGGCGAATATGATCTAT  | 3552 |
| Q | 1407 | PheValThrValProLysAspTyrValValPheGlyGlnPheAlaPhePheHisThrAla    | 1426 |
| D | 3553 | TCGATCAAGGGGCCCAAGGAATTTGTGTGTTTGGACAGTTTGGCTATTTCACAGACTGGC    | 3612 |
| Q | 1427 | LeuAsnAspValValGluValHisAspGlyHisSerGlnHisSerGlnHisSerGlnHisSer | 1446 |
| D | 3613 | CTGACACACTGGCGAGAAATTGTTGAATGAACCACTCCAGCCAGGCTTCACAGTCT        | 3672 |
| Q | 1447 | LeuSerGlySerHisThrGlyGlySerLeuProLeuAlaThrSerAsnGlnValLeuIle    | 1466 |
| D | 3673 | CTCTCTGGTTCCACTTCAGGTGAAACCTCCCGCTGCTCAATCAATCAATCAATTCGTTCT    | 3732 |
| Q | 1467 | LysPheSerAlaLysGlyLeuAlaProAlaArgGlyPheHisPheValTyrGlnAlaVal    | 1486 |
| D | 3733 | CGCTTCAGCGCAAGAGCGGAGGCTTGCACGGGGTTTCCACTTCGTTCAACCAAGCGCTC     | 3792 |
| Q | 1487 | ProArgThrSerAlaThrGlnCysSerSerValProGluProArgTyrGlyValArgLeu    | 1506 |
| D | 3793 | CCAGCGACCAAGTACACGACAGTGCAGCTCCGCTCCAGCCACGATATGGAGAAAGATT      | 3852 |
| Q | 1507 | GlySerAspPheSerValGlyAlaIleValArgPheGlyCysAsnSerGlyTyrAlaLeu    | 1526 |
| D | 3853 | GGTTCTGATGTTCTTCGAGGCTTCATGTCGATTCGATTCGAGTGCAACCAAGGTACCTGCTG  | 3912 |
| Q | 1527 | GlnGlySerProGlnIleGlyCysLeuProValProGlyAlaLeuAlaGlnTyrAsnVal    | 1546 |
| D | 3913 | CAAGGCTTCACAGCGATCCGTTGCATCTGTGCACAAAGCTTTGGCCGAGTGAATGAC       | 3972 |
| Q | 1547 | SerAlaProThrCysValValProCysGlyGlyAsnLeuThrGluArgArgGlyThrIle    | 1566 |
| D | 3973 | ACCATCCCAAGCTGTGATGTTCCATGTCAGTGGGCATTTCTCAGAGAAAGGACATC        | 4032 |
| Q | 1567 | LeuSerProGlyPheProGluProTyrLeuAsnSerLeuAsnCysValTryptylIleVal   | 1586 |
| D | 4033 | TTATCTCCAGGCTACCTCGAGCCCTATGGGAACAACCTGAACGTGTATGGAAGATCAT      | 4092 |
| Q | 1587 | ValProGluGlyAlaGlyIleGlnIleGlnValValSerPheValThrGlnGlnAsnTyr    | 1606 |
| D | 4093 | GTAATCCGAGGGCTCAGGGATCCAGATCCAAATGATATAGCTTTCACGACGACGAATCG     | 4152 |
| Q | 1607 | AspSerLeuGluValPheAspGlyAlaAspAsnThrValThrMetLeuGlySerPheSer    | 1626 |
| D | 4153 | GACTCCCTCGAGATCCATGACGGGAGACATGACGGCCCCACAGACTGGGCACTTCTCA      | 4212 |
| Q | 1627 | GlyThrThrValProAlaLeuLeuAsnSerThrSerAsnGlnLeuHisPheTyr          | 1646 |
| D | 4213 | GGTACCAACAGGCCCGGACCTGCTGAATAGAACCTCCAAACAGCTTCGCTGCACCTTCAG    | 4272 |
| Q | 1647 | SerAspIleSerValSerAlaAlaGlyPheHisIleGluGlyTyrThrValGlyLeuSer    | 1666 |
| D | 4273 | TCCGACATCAAGTGTGGCCGCTGCGGCTTTCACCTGGAATACAAACGGTGGGTCTGGCT     | 4332 |
| Q | 1667 | SerCysAspProGluProAlaValProSerAsnGlyValIleThrGlyGluArgTyrLeuVal | 1686 |
| D | 4333 | GCGTGCACAGAACTGCTCTCCGCGACAAAGCATCAAGATAGAGAACCGCTATATGATG      | 4392 |
| Q | 1687 | AsnAspValValSerPheGlnCysGluProGlyTyrAlaLeuGlnGlyHisAlaHisIle    | 1706 |
| D | 4393 | AACGATGTGCTGCTTCCAGTGCAGACCTGGGTATCACTTCGAGGGCCGCTCAACATT       | 4452 |
| Q | 1707 | SerCysMetProGlyThrValArgArgTyrAsnTyrProProProLeuCysIleAlaGln    | 1726 |
| D | 4453 | TCTTGATGCCGGGAACTGTACGTCCGTGGAATATCTCCCTCTGTGCATTGGCCACC        | 4512 |
| Q | 1727 | CysGlyGlyThrValGluGluMetGluGlyValIleLeuSerProGlyPheProGlyAsn    | 1746 |
| D | 4513 | TGTGTGGACACTGACCAACGATGAGTGGATATCCTGAGCCCAAGGCTTCCAGAGGTCA      | 4572 |
| Q | 1747 | TyrProSerMetAspCysSerTryptylIleAlaLeuProValGlyPheGlyAlaHis      | 1766 |
| D | 4573 | TACCCCAACACCTGGACTCACCTGGAAATATCCCTGGCCATTGGCTATGTGTGACAT       | 4632 |

```
OY 1767 ILeGlnPheLeuAenPheSerThrGluProAenHisAspTyrIleGlnIleArgAsnGly 1786
DB 4633 ATCCAAATTTCTGAATTTCTCACTGAAGCCAAACATAGTACTCGAGATCCAAAGTGC 4692
OY 1787 ProTyrGlnThrSerArgMetMetGlyArgPheSerGlySerGlnLeuProSerSerLeu 1806
DB 4693 CCTTACCAAGTAGTCCAAATGATGGGACAGTTACGTGGCCCTGACCTGGCTGCCTCAG 4752
OY 1807 LeuSerThrSerHisGlnThrThrValTyrPheHisSerAspHisSerGlnAenArgPro 1826
DB 4753 CTGGACCCACACATGAAACCCCTCATCGCTTCTATGTGACCACTACAGAAACCCGCAA 4812
OY 1827 GlyPheLeuLeuGlnTyrGlnAlaTyrGlnIleuGlnGluCysProAspProGluProPhe 1846
DB 4813 GGATTTAACTCACTTCCAAAGCTTATGAGTTACAGAACTGCCGGAACCCAGCCGATTC 4872
OY 1847 AlaAsnGlyIleValAlaArgGlyAlaGlyTyrAsnValGlyGlnSerValThrPheGluCys 1866
DB 4873 CAGAAAGGGTTCAATGATCAACTCCGATTACAGCGTGGCCAGTCGATCTCATTTGAGTGC 4932
OY 1867 LeuProGlyTyrGlnLeuThrGlnHisProValLeuThrCysGlnHisGlyThrAsnArg 1886
DB 4933 TACCGGGGTACATCTTGCTAGGACCCGTGTCTACCTGCAGACATGGCACTGACAG 4992
OY 1887 AsnTyrAspHisProLeuProLysCysGlyValProCysGlyGlyAsnIleThrSerSer 1906
DB 4993 AACTGAAACTAACCTTTCCACAGGTGTGAGCGTCCCTGTGGATATATGATCATACAG 5052
OY 1907 AsnGlyThrValTyrSerProGlyPheProSerProTyrSerSerGlnAspCysVal 1926
DB 5053 AATGGCACCATTTATTCCTCGGGTTCCTCCAGACAGATCAATTCCTGAAGAGCTCCGTC 5112
OY 1927 TyrLeuIleThrValProIleGlyHisGlyValArgLeuAsnLeuSerLeuGlnThr 1946
DB 5113 TGCTGTGCTCTGCTCCTCCAGACATGAGTGAATCACTTCACTTGCAGACT 5172
OY 1947 GluProSerGlyAspPheIleThrIleTyrAspGlyProGlnGlnThrAlaProArgLeu 1966
DB 5173 GAGGCTGTAAATGATCATCATCGCTGTGGATGGTCTTGACCAAGAACTCGCTCAGCTC 5232
OY 1967 GlyValPheThrArgSerMetAlaLysThrValGlnSerSerAsnGlnValLeu 1986
DB 5233 GGGGTCTTCACTGAAACACTGCTCCGAGACAGCATACACCTCCCAACAGGCTCTTG 5292
OY 1987 LeuLysPheHisArgAspAlaIleThrGlyIlePheAlaIleAspSerAlaTyr 2006
DB 5293 CTCAATTCACAGCGATTTCTCCATGAGGCTTCTTGTCTCAATTTTCATGCAATT 5352
OY 2007 ProLeuThrLysCysProProProThrIleLeuProAsnAlaGlyValValThrGluAen 2026
DB 5353 CAACCTAAGAGGTGCCCGCTCTCCAGTAGTCCGACAGCTGACCTGTCACAGAAAT 5412
OY 2027 GluGluPheAsnIleGlyAspIleValArgTyrArgCysLeuProGlyPheThrLeuVal 2046
DB 5413 GAAAGCTTTGAATTAAGGAGCTTGTAAAGTACAGTGCATCCAGAGGTACACGCTGTG 5472
OY 2047 GlyAsnGluIleLeuThrCysLysLeuGlyThrTyrLeuGlnPheGlnGlyProPro 2066
DB 5473 GGAAGTGAACCCCTGACATGCAAGCTCAGCTCAGCTATTTGTTCCAAAGGCTCTCAGCT 5532
OY 2067 IleCysGlyValHisCysProThrAsnGluLeuLeuThrAspSerThrGlyValIleLeu 2086
DB 5533 ACCTGTGAAGCAACAAAGCCAGCAATGAAAGTCCGAACAGAGCTTCTGGGTGATTCTC 5592
OY 2087 SerGlnSerTyrProGlySerTyrProGlnPheGlnThrCysSerThrLeuValArgVal 2106
DB 5593 AGTCTCGGTACCCAGGACACATTTTAACTCCCAACATGTGTGAGTAAAGG 5652
OY 2107 GluProAspTyrAsnIleSerLeuThrValGluTyrPheLeuSerGlyLysGlnTyrAsp 2126
DB 5653 GAGCCAAACTTAACTTACGCTTGTGACACCTTTCAAGTGAAGAAAGCAATTTGAT 5712
OY 2127 GluPheGlnIlePheAspGlyProSerGlyGlnSerProLeuLeuLysValLeuSerGly 2146
DB 5713 GCACGTGAAGTATTTGATGATGTTCTTGGGCAAGCTCTTGTGTGATGCTTAAAGTGG 5772
OY 2147 AsnTyrSerAlaProLeuIleValThrSerSerSerAsnSerValTyrLeuArgTyrPsr 2166
DB 5773 AACCACTGAACAGTCCAAATTTTACACAGAAATACATATCTTACCTCGCTGTGC 5832
OY 2167 SerAspHisAlaTyrAsnArgGlyGlyPheLysIleArgTyrSerAlaProTyrCysSer 2186
DB 5833 ACAGATCATGACACACGAAGAAAGATTCAAGATTGCTATGACGCTCTTACGCAAC 5892
OY 2187 LeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThrSerThrGlnProGlyGly 2206
DB 5893 CTCACCTTACACTCAAGATATGCTGGCTTTTAAATTAACCGACGGGCCCTGGAGGAC 5952
OY 2207 SerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGlyHisSerMetAlaIleCys 2226
DB 5953 AAGTGCAGATATTTCTGCAAGCTGTGATATGATATGGCCACAGAACCCACCTGC 6012
OY 2227 ThrArgHisProGlnGlyTyrHisLeuThrPserGluAlaIleProLeuCysGlnAlaLeu 2246
DB 6013 AGCGGAAACCCAGTGGCGGTGACAGTGGGACCTGATGGACCCGCTTGGCAGGCTGC 6072
OY 2247 SerCysGlyLeuProGluAlaProLysAsnGlyMetValPheGlyLysGluTyrThrVal 2266
DB 6073 TCTGTGAAATTTCCAGAGCTTCAGAAATGGCTCTTCAACAGCAATGATGATCACTTA 6132
OY 2267 GlyThrLysAlaValTyrSerCysSerGlnGlyTyrHisLeuGlnAlaGlyValAlaGluAla 2286
DB 6133 GACAGTAAAGTCACTTATGATATGATGAAGCTTCAAGCTGATGACAGTACAGAACCC 6192
OY 2287 ThrAlaGluCysLeuAspThrGlyLeuThrPserAsnArgAsnValProProGlnCysVal 2306
DB 6193 ACTGTGTGTGTCAGAAAGATGGCTGTGGAGCAACAGAGAAAGCCACCACTGGCAA 6252
OY 2307 ProValThrCysProAspValSerSerIleSerValGlnHisGlyArgTyrArgLeuLeu 2326
DB 6253 CCGGTGCCCTCCAGCATGAAAGGCCAGCTGCAGAGCACGTGCTTGAAGGCTGGTT 6312
OY 2327 PheGluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyrTyr 2346
DB 6313 TCGGATCATGTAATGATATGAGCTCAAGTCTCTCAGCTGATAGTCTGGCTACTTC 6372
OY 2347 TyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLysTyrPserLeuGlyAspSer 2366
DB 6373 TTGAGGCTCAGAGGCTGTTCAGTGCACAGCCCAATGGGACCTGGAACATGAGAGC 6432
OY 2367 ThrProThrCysArgIleIleSerCysGlyGluLeuProIleProProAsnGlyHisArg 2386
DB 6433 AGACCAAGATGTAAGTATCTCTGTGGAAGCTGTCTTCCCCAAATGATGACAG 6492
OY 2387 IleGlyThrLeuSerValTyrGlyAlaThrAlaIlePheSerCysAsnSerGlyTyrThr 2406
DB 6493 ATAGGAGCTCACTATGATGAGGACCAACCCCTTCTTCACTCAATAACGGCTACACA 6552
OY 2407 LeuValGlySerArgValArgGluCysMetAlaAsnGlyLeuThrPserGlySerGlyVal 2426
DB 6553 CTGTAGGCTCCCATGTCGGAGAGCTGTGGCCATGATGCTCTGAGCGGATTCGAACA 6612
OY 2427 ArgCysLeuAlaGlyHisCysGlyThrProGluProIleValAlaAsnGlyHisIleAsnGly 2446
DB 6613 AGTGTCCGTGGGGTCACTTGTGCTCCAGAACCCATTTGATGATGCAATAGTGC 6672
OY 2447 GluAsnTyrSerTyrArgGlySerValValTyrGlnCysAsnAlaGlyPheArgLeuIle 2466
DB 6673 GATGCTTCACTACAGAGACACAGTGTCTACCAAGCAACCTGGGTTCGACTGTA 6732
OY 2467 GlyMetSerValArgGlyCysGlnGlnAspHisHisThrPserGlyLysThrProPheCys 2486
DB 6733 GGCAGCTGTGAGGATTTGCTGACAGACCAACAGTGTGGGGGACACCCCGCTTGC 6792
OY 2487 ValProIleThrCysGlyHisProGlyAsnProValAsnGlyLeuThrGlnGlyAsnGln 2506
```



(INCY-) INCYTE GENOMICS INC.

PA La1 PG, Baughn MR, Yao MG, Wajia NK, Elliot VS, Xu Y;  
 XX Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM,  
 PI Hatalla AA, Ghadi AR, Thangavelu K, Sanjanwala MM, Tang YT,  
 PI Rakhumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK,  
 PI Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;  
 XX WPI: 2002-713453/77.  
 DR P-PSDB: MA019415.

PT New human molecules for disease detection and treatment (MDT), useful  
 PT for diagnosing, treating and preventing diseases or conditions associated  
 PT with the aberrant MDT expression, e.g. cancer, AIDS, asthma, diabetes,  
 PT hepatitis.

PS Claim 11; Page 174-176; 177pp; English.

XX The present invention relates to human proteins and coding sequences of  
 CC molecules for disease detection and treatment MDT. The sequences can be  
 CC used in the treatment of diseases associated with the decreased  
 CC expression or overexpression of MDT, such as cell proliferative (cancer,  
 CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,  
 CC allergies, Addison's disease, asthma), developmental (dwarfism, renal  
 CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,  
 CC epilepsy) and cardiovascular (congestive heart failure, myocardial  
 CC infarction, angina pectoris) disorders. The present sequence is a coding  
 CC sequence of the invention

SQ Sequence 6004 BP; 1272 A; 1795 C; 1615 G; 1322 T; 0 U; 0 Other;

Alignment Scores:

| Pred. No.:             | 0       | Length:       | 6004 |
|------------------------|---------|---------------|------|
| Score:                 | 9077.50 | Matches:      | 1720 |
| Percent Similarity:    | 86.47%  | Conservative: | 5    |
| Best Local Similarity: | 86.22%  | Mismatches:   | 8    |
| Query Match:           | 53.44%  | Indels:       | 265  |
| DB:                    | 6       | Gaps:         | 8    |

US-10-016-248-2 (1-3104) x AAU49944 (1-6004)

QY 859 AsnSerValThrGlyThrGlnGlyThrLeuLeuSerProAsnPhleProValAsnTyrAsn 878  
 DB 75 AATTCAGTCAAGACTCA--GGACTTG-CTGTCCCAAGCT-CCCGGAATTCACAT 129

QY 879 AsnAsnHisGluCysIleTyrSerIleGlnThrGln-Pro-GlyTyrGlyIleGlnLeu 898  
 DB 130 GACAAATCCGTAATGCAT-TACTCCATGCAAGCCAGTCCAGGTAAGCCAAATTCAGCTCA 188

QY 898 ValAlaArgAlaPheGluLeuSerGluGlyAspValLeuValTyrAspGlyAsnAsn 918  
 DB 189 AAGCCAGGGCATTCGAACTCTCCGAAGAGATGCTCCCAAGGTTTGTATGGCAACACAA 248

QY 918 snSerAlaArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAsn 938  
 DB 249 ACTCCGCCCGTTGGCTGGAGAGTTTAAAGCATTTCTAAGATGAGGGGGTGAAGCTTGAACA 308

QY 938 erThiSerSerSerLeuTyrPleuAspPheIleThrAspAlaGluAsnThrSerIleGlyP 958  
 DB 309 GCACATCCAGCAGCTGTGGCTGATTTTCATCATCTGATAAACAACACAGAGAGGCT 368

QY 958 heGluLeuHisPheSerSerPheGluLeuIleGlyCysGluAspProGlyThrProLysP 978  
 DB 369 TTGAATCGACATTTTCCAGCTTGAACATCAAAAGTGAAGACCCAGAACCCCAAGT 428

QY 978 heGlyTyrIleValHisAspGluGlyValPheAlaGlySerValSerPheSerCysA 998  
 DB 429 TTGGCTCAAGGTTCTATGATGAAGTCATTTTTCAGAGAGCTTCGTTCTTCAAGTGG 488

QY 998 spProGlyTyrSerLeuArgGlySerGluGluLeuLeuCysLeuSerGlyGluArgTyr 1018  
 DB 489 ACCCTGGATACAGCTCGGGGGTGAAGAGAGCTGCTGTCTGATGGAGAGCCCGGGA 548

QY 1018 hTrpAspArgProLeuProThrCysValAlaGluCysGlyGlyThrValAlaGlyGlyVal 1038  
 DB 549 CCGGAGACCGGCTCTGCCACCTGTGTGCCAGAGTGGAGAGACAGAGAGAGAGAGAG 608

QY 1038 alSerGlyGlnValLeuSerProGlyTyrProAlaProTyrGlnHisAsnLeuAsnCysI 1058  
 DB 609 TGTCCGGGGAGAGTGTGTATCAACCCGGGTATTCAGCTCCCTTAATACCAATTCACATTCGA 668

QY 1058 leTrpThrIleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspT 1078  
 DB 669 TCTGGACCATTCAG 728

QY 1078 heGluGluValHisAspValLeuArgGlyIleThrAspGlyProValGluSerGlyValLeu 1098  
 DB 729 CAGAGAGAGTCAAGAGAGTGTGGCATGTGGATGGAGATGGAGATGGAGATGGAGATGGAG 788

QY 1098 euLysGluLeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerVal 1118  
 DB 789 TGAAGAGAGTGAAGTGGCCGCGCCGCTGCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 848

QY 1118 alLeuGlnPheSerThrAspPhePheThrSerIleGlyPheAlaIleGlnPheSerV 1138  
 DB 849 TCTTCAGATTCAAGCACTGACTTCTTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 908

QY 1138 alSerThrAlaThrSerCysAsnAspProGlyIleProGlnAsnGlySerArgSerGly 1158  
 DB 909 TGTCCAG 968

QY 1158 spSerTyrPleuAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeuGln 1178  
 DB 969 ACAGTTGGAG 1028

QY 1178 lYserAlaGlnIleSerCysValIleGluAsnArgPhePheThrGlnProSerProP 1198  
 DB 1029 GAAAGTCAAGAGATCACTGTGTGAAGATCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1088

QY 1198 roThrCysIleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerP 1218  
 DB 1089 CAACATGATGAGTGTCTCCGCGGGGAGAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1148

QY 1218 roAsnTyrProGluProTyrProProGlyIleGlyCysAspTyrIleValThrValSerP 1238  
 DB 1149 CAATATACCAAG 1208

QY 1238 roAspTyrValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheL 1258  
 DB 1209 CAGACTACGATGAGCTGCTGTGTATTAACATCTTAACATGAGAGAGAGAGAGAGAGAGAG 1268

QY 1258 euHisIleTyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerG 1278  
 DB 1269 TCCATATCTACAG 1328

QY 1278 lLeuPProGlyValIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspA 1298  
 DB 1329 AGCTCCAG 1388

QY 1298 lAsSerValSerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGlyLeuSerCysP 1318  
 DB 1389 CATCTGTAG 1448

QY 1318 heAspProGlySerIleLeuAsnGlyThrArgValGlySerAspLeuLeuGlySerS 1338  
 DB 1449 TTGATCTCGTTCCATCAAG 1508

QY 1338 erValThrTyrTyrCysHisGlyGlyTyrGluValIleGlyThrSerThrLeuSerCysI 1358  
 DB 1509 CCGTACCTTACTATCTGACAG 1568

QY 1358 lLeuGlyProAspGlyLysProValTyrAsnAsnProAspProValCysThrAlaProC 1378  
 DB 1569 TCCGTGGGGCCGTAG 1628

QY 1378 ysGlyGlyGlnTyrValGlySerAspGlyValValLeuSerProAsnTyrProGlnAsnT 1398

```

Db      1629 GTGGGGGACGATAGTGGGTTCCGACGAGTGTCTTGTCCGCCCACTAACCCCGAGAACT 1688
Qy      1398 YRTHRSerGlyGlnIleCysLeuTYRProValProAlaLeuLysSerThrSera 1418
Db      1689 ACACCAAGTGGACAGATCTGCTGTGATTTTGTACTGTGCCCCAGAGACTATGTGTGTTTG 1748
Qy      1418 LysIlePheAlaPhePheIleThrAlaLeuAsnAspValValGluValHisAspGlyHis 1438
Db      1749 GCCAGTTGCGCTTCTTTTCACAGGCCCTCAACAGAGTGTGAGGTTTCACAGCGACACA 1808
Qy      1438 ergIHisSerArgLeuLeuSerSerLeuSerGlySerHisIleThr----- 1452
Db      1809 GCCAGCACTCGCGGCTCTCAAGCTCCCTCTGGGCTCCCACTACAGGATCCGGGGCTCGG 1868
Qy      1452 ----- 1452
Db      1869 CCAGTGTGGAGTGTGTGTGGGCGGGGGCATCAAGTCCGGCTAAAGAGAGAGGCTCTA 1928
Qy      1452 ----- 1452
Db      1929 GAAGACCCCACTGGCGCCGAGGTGGAACCTTACGGCTCTGCGCTGTCTGTCTGTG 1988
Qy      1452 ----- 1452
Db      1989 CTGTGTCAACAAGCAGCAGCAGCGCTCGTAGAGCTCCAACTAGCGGGGCTTCAGCAGCT 2048
Qy      1452 ----- 1452
Db      2049 GCCCTCAACCAAGACTGTGTCTACACCGCCCTTGTGTGTAGCTTCTCTGTTGAATG 2108
Qy      1452 ----- 1452
Db      2109 GCAACTACACTAATTGGCTGCAAGTCCAAGTTGTGTCTCTCTCCCTGGCCCATCTGTA 2168
Qy      1452 ----- 1452
Db      2169 CTGCACCAAGCAGAAAGATATACCTTTGTCTCTGCTACAAAGCTGTCACTACCTCTG 2228
Qy      1453 -----GlyIleSerLeuProLeuAlaThrSerAsnGlnValLeuIle 1467
Db      2229 TTTCCTGTGCCCATGACGAGAAATCACTGCTTGGCCCACTCCAACTCAATCTCTCATTA 2288
Qy      1467 YSPheSerAlaIleValGlyLeuValProAlaArgGlyPheHisPheValTYRgln----- 1484
Db      2289 AGTTCAAGGCGAAAGGCTCGCACACGACGAGGAGCTTCACTTGTCTACCAAGGTATGG 2348
Qy      1485 -----AlaValProArgThrSerAlaThrGlnCysSerSerValPro 1499
Db      2349 AGGACATGAGACGCCGAGCGGTTCTCTGAAACCAAGCCAGCGACGTGACGTGTGCCG 2408
Qy      1499 LuProArgTYRgLYLysArgLeuGlySerAspPheSerValGlyAlaIleValArgPheG 1519
Db      2409 AACCCCTCTATGGCAGAGGCTGGGAGTAGTCTCTGGTGGGGGCAATCGCTCGCTTGG 2468
Qy      1519 LuCysAsnSerGlyTYRAlaLeuGlnGlySerProGluIleGluCysLeuProValPro 1559
Db      2469 AATGCAACTCGGCTATAGCCTGTGAGGGGTGCGCAGAGATGAGATGCTCTCTGTGCTG 2528
Qy      1539 LysAlaLeuAlaGlnTrpAsnValSerAlaProThrCysValValProCysGlyValAsn 1559
Db      2529 GGGCTTGGCCCATATGAAATCTTCAGCGCCCACTGTGTGTGTCCTGTGTGAGGACAAC 2588
Qy      1559 eutHrGluArgArgGlyThrIleLeuSerProGlyPheProGluProTYRLeuAsnSer 1579
Db      2589 TCACAGAGCGCAGGGGACCATCTCTGTCTCCCTGGCTTCCAGAGCCCTTACCTCAACGCC 2648
Qy      1579 euaenCysValTrpLysIleValValProGluGlyAlaGlyIleGlnIleGlnValValS 1599
Db      2649 TCAACTGTGTGAGAAATGTGTGTCTCCCGAAGGCGCTGCGATCAAGATTCAAATTGTCA 2708
Qy      1599 erPheValThrGluGlnIleAsnTrpAspSerLeuGluValPheAspGlyValAspAsnThrV 1619

```

---

```

Db      2709 GTTTTGTACAGACAGAACTGGGACTCGCTGGAAGTATTGATGTGACAGATAACTG 2768
Qy      1619 alThrMetLeuGlySerPheSerGlyThrThrValProAlaLeuLysSerThrSera 1639
Db      2769 TAAACATGTGGGGAGTTTCTCAGAGAACAAACCGGCTCGCTTCTGAAACGACCTCCA 2828
Qy      1639 snGlnLeuTYRLeuHisPheTYRSerAspIleSerValSerAlaAlaGlyPheHisLeuG 1659
Db      2829 ACCAGCTTCACTTCAATTTCTTCAATATACAGATATCTGACGTGGCTTCCACTTGG 2888
Qy      1659 LuTYRLeuThrValGlyLeuSerSerCysProGluProAlaValProSerAsnGlyVal 1679
Db      2889 AGTACAAACAGGTGGGCTGTGAGGATGTTGCCGGAACCTGCTGTGCCCCAGTACGGGGTGA 2948
Qy      1679 YRTHRGlyGluArgTYRLeuValAsnAspValValSerPheGlnCysGluProGlyTYR 1699
Db      2949 AGACTGGCAGCGCTACTGTGTGATGATGTGTGTCTTTCAGTGTGAGCCGGGATATG 3008
Qy      1699 LysLeuGlnGlyHisAlaHisIleSerCysMetProGlyThrValArgArgTrpAsnTYR 1719
Db      3009 CCTTCAGGGCCAGCCCACTCTCTGATGCCCCGGAACAATGCGGCGATGGAATTAAC 3068
Qy      1719 roProProLeuCysIleAlaGlnCysGlyTYRThrValGluGluMetGluGlyValIle 1739
Db      3069 CTCTCCACTCTGTATTGACAAGTGTGGGGAACAGTGGAGAGATGAGGGGGTGTATCC 3128
Qy      1739 euserProGlyPheProGlyAsnTYRProSerAsnMetAspCysSerTrpLysIleAla 1759
Db      3129 TGAGCCCCGGCTTCCAGGCAACTACCCAGTAAACATGAGACTGCTCTCGAATAATACAC 3188
Qy      1759 eupProValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrGluProAsnHis 1779
Db      3189 TGCCCGTGGGCTTTGAGCTCAATCCAGTCTCTGAATCTTCTCCACCGAGCCCAACACG 3248
Qy      1779 sPTYRileGluIleArgAsnGlyProTYRglnThrSerArgMetMetGlyArgPheSerg 1799
Db      3249 ACTACATAGAAATCCGGAATGGCCCTATAGACACAGCCCATGATGGAAAGATTACGTG 3308
Qy      1799 LysArgIleLeuProSerSerLeuLeuSerThrSerHisGlyIleThrValTYRPhHis 1819
Db      3309 GAGCGAGCTTCCAAAGCTCTCTCTCTCCACGTCACGAGACACACCGGTATTTCCACA 3368
Qy      1819 erAspHisSerGlnAsnArgProGlyPheLysLeuGluTYRgln----- 1833
Db      3369 GCGACCACTCCAGAAATCGCGCAGATTCAGGCTGAGATTCAGAGATTGACTTACTCCC 3428
Qy      1833 ----- 1833
Db      3429 ACCAGATTCTTCTCTCTGAGAGTTTGAATCTCTCGGAGTTGGAAGAACAACACTCA 3488
Qy      1834 -----AlaTYRgln 1837
Db      3489 CTCTCCCGCGCGCGCTTCTATGTCTGAGATCTTGATCTCTGTGTGGAAGCCATGAA 3548
Qy      1837 eugGlnGlyCysProAspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTYR 1857
Db      3549 TTCAAGAGTCCCAAGACAGAGCCCTTTCACATGAGCATTTGAGGGAGAGCTGGCTACA 3608
Qy      1857 snValGlyIleSerValThrPheGlyCysLeuProGlyTYRglnLeuThrGlyHisIlePro 1877
Db      3609 ACGTGGGACATACATGATCTTGAAGTCTCTCCGGGGATTCATATGACGTGGCAACCTG 3668
Qy      1877 alLeuThrCysGlnHisGlyThrAsnArgAsnTrpAspHisProLeuProLysCysGlu 1897
Db      3669 TCTCTACGTGTCAACATGACCAACACCGGAATCGGACCAACCCCTGCCCCAAGGTGTAA 3728
Qy      1897 alProCysGlyValAsnIleThrSerSerAsnGlyThrValTYRserProGlyPhePro 1917
Db      3729 TCCCTTGTGGCGGAACATACATTTCTTCAACGACATGTGTACTCCCGGGGTCTCTTA 3788
Qy      1917 erProTYRserSerSerGlnAspCysValTYRLeuIleThrVal--ProIleGlyHis-G 1936
Db      3789 GCCCGTACTCAGACTCCCAAGAGCTGTGTCTGGCTGATCACCGTGGCCAAATTGGCCATG 3848

```

QY 1316 |yVal|ArgLeuAsnLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleT 1956  
 Db 3849 GCGTCCGCTCAACCTGAGCTGCTGGAGACAGAGCCCTCTGGAGATTTCATCCACTCT 3908  
 QY 1956 |rPaapGlyProGlnGlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLysL 1976  
 Db 3909 GGGATGGGGCCACACCAACAGCACCGGCTCGGCGTCTTCACCCGAGCATGGCCACA 3968  
 QY 1976 |yEThrValGlnSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 1996  
 Db 3969 AAACAGTCAGAGTTCATCCACAGAGTCTGCTCAAGTTCACCGGATGACAGCCAG 4028  
 QY 1996 |yGlyIlePheAlaIleAlaPheSerAlaTyProLeuThrLysCysProProThrT 2016  
 Db 4029 GGGGAGTCTTCGACATAGCTTCTCCGCTTATCCTACCAATGCTCTCTCCACCA 4088  
 QY 2016 |LeuProAsnAlaGluValValThrGluAsnGluGluPheAsnIleGlyAspIleValA 2036  
 Db 4089 TCTTCCCAAGCGGAGAGTGTCTCAGAGATGAAATTCATATAGGTGACATGCTAC 4148  
 QY 2036 |rGlyArgCysLeuProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuG 2056  
 Db 4149 GCTACAGATGCTCTCCGCTTTCCTTACCTAGTGGGAAATGAATTCGACCTGCAACTTG 4208  
 QY 2056 |yThrTyLeuGlnPheGluGlyProProProIleCysGluValHisCysProThrAsnG 2076  
 Db 4209 GAACCTACCTGCGATTGGAAGAGACACCCCGATATGTGAAGTGCATGCTCCCAAAAG 4268  
 QY 2076 |LeuLeuThrAspSerThrGlyValIleLeuSerGlnSerTyProProGlySerTyProG 2096  
 Db 4269 AGCTTCTGACAGACTCAGAGGGGTATCTGAGCCAGAGTACCTCGGAAGGTATACCCC 4328  
 QY 2096 |InPheGlnThrCysSerTrpLeuValArgValGluProAspTyAsnIleSerLeuThrV 2116  
 Db 4329 AGTTCCAGACCTGCTGTGGCTGTGAGTGAAGTGGAGCCCATTAATCTCCCTCAGAG 4388  
 QY 2116 |AlGluTyRheLeuSerGluGluGlnTyRAspGluPheGluIlePheAspGlyProSerG 2136  
 Db 4389 TGAAGTACTTCTCAGAGAGAACCAATATGATGATTGATGATTGATGCTCATACAG 4448  
 QY 2136 |yGlnSerProLeuLeuValAlaLeuSerGlyAsnTyRserAlaProAlaIleValThs 2156  
 Db 4449 GACAGAGTCTCTGCTGAAGACCCCTCAGTGGGAATTAATCACTGCTCCCTGATGTGACA 4508  
 QY 2156 |eSerSerAsnSerValTyRLeuArgTyRSerSerAspHisAlaTyRAsnArgLysGlyP 2176  
 Db 4509 GCTCAAGCAACTCTGTGTACTGCTGTTGGTCACTGATCAAGCTTCAATCCGAAGGGCT 4568  
 QY 2176 |heLysIleArgTyRserAlaProTyCysSerLeuProArgAlaProLeuHisGlyPheT 2196  
 Db 4569 TCAAGATCCGCTATTC----- 4585  
 QY 2196 |LeuGlnGlnThrSerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyT 2216  
 Db 4586 -----GGCCAGACACACCCAGCCCGGGGGCTCCATCCACTTTGGCTCAACCGCGGCT 4640  
 QY 2216 |yRArgLeuValGlyHisSerMetAlaIleCysThrArgHisProGlnGlyTyRHisL 2236  
 Db 4641 ACCGCTGTGGGACACACGATGCGCATCTGTAACCGGCAACCCGAGGCTACCACTGT 4700  
 QY 2236 |rSerSerGlnAlaIleProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProLysA 2256  
 Db 4701 GAGACCAAGCCATCCCTCTCTGTCAAGCTCTTTCCTGTGGGCTTCTCGTGGGCCCCCAAGA 4760  
 QY 2256 |snGlyMetValPheGlyLysGluTyRThrValGlyTyRThrLysAlaValTyRserCysSerG 2276  
 Db 4761 ATGGAATGTGTGTTGGCAGAGTACACAGTGGGAACCAAGCCATGTAAGCTCACTGAGT 4820  
 QY 2276 |LugIlyTyRHisLeuGlnAlaGlyAlaGluAlaThrAlaGluCysLeuAspThrGlyLeuT 2296  
 Db 4821 AAGGCTACCACTCCAGGAGGCGGTGAGGCACTGCAAGATGTCTGGACACAGGCTCAT 4880

QY 2296 |rPserAsnArgAsnValProProGlnCysVal-----P 2307  
 Db 4881 GAGACACCGCATGTCTCCACACAGTGTGTCCGTGATCTTCGGGCAATGAGCGGGT 4940  
 QY 2307 |roValThrCysProAspValSerSerIleSerValGluHisGlyArgTyRPaArgLeuIleP 2327  
 Db 4941 CTGTGACTGTCTGTATGTCAATAGATCAGGTGGAGATGAGCCGATGAGAGCTTATCT 5000  
 QY 2327 |heGluThrGlnTyRLeuPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyRTyT 2347  
 Db 5001 TTGAGACACAGTATCAGTTCAGGCGCCAGCTGATCTCATCTGTATCCTCTGCTACT 5060  
 QY 2347 |yRThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLysTrpSerLeuGlyAspSerT 2367  
 Db 5061 ATACTGGCCAAAGGTATCTCCGTCTGAGCCCAATGGCAATGAGACTCGGAGACTTGA 5120  
 QY 2367 |hProThrCysArgIleIleSerCysGlyLeuProIleProProAsnGlyHisArgI 2387  
 Db 5121 CCCCCACCTGCGGAATCATCTCTGTGAGAGCTCCCGATTCCCCCAATGGCCACCGCA 5180  
 QY 2387 |LeGlyThrLeuSerValTyRValAlaThrAlaIlePheSerCysAsnSerGlyTyRThrL 2407  
 Db 5181 TGGAACTACTGTCTATCGGGGCAACAGCCATCTTCTGCAATTCGGATACACAC 5240  
 QY 2407 |euValGlySerArgValArgGluCysMetAlaAsnGlyLeuTrpSerGlySerGluValA 2427  
 Db 5241 TGTGGGGCTCCAGGGTGGGTGATGTCAAGCCAAATGGGCTCTGAGTGGCTTGAAATTC 5300  
 QY 2427 |rGlyLeu----- 2429  
 Db 5301 GCTGCTTGGCACACTGACCAAGCTCAGTCCATTTCTATAGCTCCTCTTCATGATGAC 5360  
 QY 2430 -----AlaGlyHisCysGlyThrProGluProIleValAsnG 2442  
 Db 5361 TCTCTTCCCATCTCCCTCAACAAAGTGGACAGTGTGGACCTCTGAGCCCATTTGCAAG 5420  
 QY 2442 |yHisIleAsnGlyGluAsnTyRserTyRArgLysSerValTyRGlncysAsnAlaG 2462  
 Db 5421 GACACATCAATGGGGAGAACTACAGCTACCGGGGAGTGTGTATCCAAATGCATGCTG 5480  
 QY 2462 |yPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisHisTrpSerGlyL 2482  
 Db 5481 GCTTCCGCTGATGGAGATGTGTGTGTGCGCATCTGCACAGAGATCATCATGCTGGGCA 5540  
 QY 2482 |yThrProPheCysVal----- 2487  
 Db 5541 AGACCCCTTCTGTGTGATGTAAAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5600  
 QY 2488 -----ProIleThrCysGlyHisProGlyAsnP 2497  
 Db 5601 ATGATGATGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 5660  
 QY 2497 |roValAsnGlyLeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValLysPheValC 2517  
 Db 5661 CTGTCAACGGCTCACTCAGGGTAACAGTTTAACTTCAAGAGATGTGTCAAGTTTGT 5720  
 QY 2517 |yAsnProGlyTyRMetAlaGluGlyAlaAlaArgSerGlnCysLeuAlaSerGlyGlnT 2537  
 Db 5721 GCAACCTGGGTATATGCTGAGGGGCTGTGATGCTCCAAATGCTGCGCAGCGGGCAAT 5780  
 QY 2537 |rPserAspMetLeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGluA 2557  
 Db 5781 GGAATGACATGCTGCCACCTGCGAATATCAACTATACAGATCTGAGACACCAAGAAA 5840  
 QY 2557 |snSerValArgGlnValHisAlaSerGlyProHisArgPheSerPheGlyThrThrValS 2577  
 Db 5841 ATATGTTCTCTCAAGTCAAGCCAGCGGCCCGGACAGATTCAAGTTTGGGACCACTGTGT 5900  
 QY 2577 |eTyRArgCysAsnHisGlyPheTyRLeuLeuGlnTyRProValLeuSerCysGlnGlyA 2597  
 Db 5901 CTTACCGGTGCAACACAGGCTTCACTCTGCGGACCCCAAGTGTCAAGTGTGCAAGGAG 5960  
 QY 2597 |spGlyThrTrpAspArgProArgProGlnCysLeu 2608

Db 5961 ATGGCACATGGGACCGTCCCCCGCCAGTGTCTC 5995

Search completed: October 18, 2004, 16:07:48  
Job time : 4182 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 09:58:21 ; Search time 150 Seconds  
(without alignments)  
11906.419 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985

Sequence: 1 MAGAPPALLPCLISDCC.....RSGVDPSTLPGRSPK 3104

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0 %  
Maximum Match 100 %  
Listing first 45 summaries

Database :

1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score   | Query Match | Length | DB ID        | Description        |
|------------|---------|-------------|--------|--------------|--------------------|
| 1          | 15364.5 | 90.5        | 3487   | 1 CSM2 HUMAN | Q7408 homo sapien  |
| 2          | 11372.5 | 67.0        | 3565   | 1 CSM1 HUMAN | Q96P27 homo sapien |
| 3          | 11305   | 66.6        | 3564   | 1 CSM1 MOUSE | Q92313 mus musculu |
| 4          | 10591.5 | 62.4        | 3670   | 1 CSM3 HUMAN | Q7407 homo sapien  |
| 5          | 8997.5  | 53.0        | 2796   | 1 CSM3 MOUSE | Q80C79 mus musculu |
| 6          | 6180.5  | 36.4        | 1466   | 2 Q72241     | Q72241 brachydanio |
| 7          | 1864.5  | 11.0        | 3567   | 2 Q96S77     | Q96S77 mus musculu |
| 8          | 1550    | 9.1         | 2489   | 2 Q16744     | Q16744 homo sapien |
| 9          | 1473    | 8.7         | 3494   | 2 Q7LC53     | Q7LC53 homo sapien |
| 10         | 1473    | 8.7         | 3623   | 2 Q60494     | Q60494 homo sapien |
| 11         | 1457.5  | 8.6         | 3620   | 2 Q9TUS3     | Q9TUS3 canis famli |
| 12         | 1416    | 8.3         | 1441   | 2 Q723G3     | Q723G3 homo sapien |
| 13         | 1415    | 8.3         | 1497   | 2 Q8BMT9     | Q8BMT9 homo sapien |
| 14         | 1404.5  | 8.3         | 3623   | 2 Q70244     | Q70244 rattus norv |
| 15         | 1390    | 8.2         | 2014   | 2 Q29530     | Q29530 pan troglod |
| 16         | 1372    | 8.1         | 1323   | 2 Q72387     | Q72387 homo sapien |
| 17         | 1371    | 8.1         | 2039   | 1 CR1 HUMAN  | P17927 homo sapien |
| 18         | 1298    | 7.6         | 1911   | 2 Q29528     | Q29528 papio hama  |
| 19         | 1190    | 7.0         | 3646   | 2 Q70737     | Q70737 anophelae g |
| 20         | 968.5   | 5.7         | 3687   | 2 Q9W333     | Q9W333 drosophila  |
| 21         | 906.5   | 5.3         | 1025   | 1 CR2 MOUSE  | P19070 mus musculu |
| 22         | 906.5   | 5.3         | 1032   | 2 Q9DC83     | Q9DC83 mus musculu |
| 23         | 906     | 5.2         | 1174   | 2 Q9VYR4     | Q9VYR4 drosophila  |
| 24         | 889.5   | 5.2         | 1033   | 1 CR2 HUMAN  | P20023 homo sapien |
| 25         | 878.5   | 5.2         | 1135   | 2 Q70137     | Q70137 anophelae g |
| 26         | 864.5   | 5.1         | 974    | 2 P91658     | P91658 drosophila  |
| 27         | 832     | 4.9         | 1045   | 2 Q46545     | Q46545 ovis aries  |
| 28         | 797     | 4.7         | 869    | 2 Q922H0     | Q922H0 mus musculu |
| 29         | 793     | 4.7         | 869    | 2 Q8NB57     | Q8NB57 homo sapien |
| 30         | 793     | 4.7         | 972    | 2 Q8ND50     | Q8ND50 homo sapien |
| 31         | 788     | 4.6         | 996    | 2 Q8TD25     | Q8TD25 homo sapien |

|    |       |     |      |              |                    |
|----|-------|-----|------|--------------|--------------------|
| 32 | 785.5 | 4.6 | 1022 | 1 TLD BRARE  | Q57460 brachydanio |
| 33 | 772.5 | 4.5 | 1008 | 2 Q9DER7     | Q9DER7 gallus gall |
| 34 | 771.5 | 4.5 | 991  | 2 Q7TSK2     | Q7TSK2 mus musculu |
| 35 | 771   | 4.5 | 1012 | 2 Q9WVW6     | Q9WVW6 mus musculu |
| 36 | 769.5 | 4.5 | 1013 | 2 Q9NOS4     | Q9NOS4 homo sapien |
| 37 | 768.5 | 4.5 | 1013 | 2 Q43897     | Q43897 homo sapien |
| 38 | 764   | 4.5 | 997  | 2 Q62223     | Q62223 mus musculu |
| 39 | 763.5 | 4.5 | 977  | 2 Q62269     | Q62269 mus musculu |
| 40 | 763   | 4.5 | 775  | 2 Q6P550     | Q6P550 mus musculu |
| 41 | 763   | 4.5 | 775  | 2 AAH63079   | AAH63079 mus muscu |
| 42 | 763   | 4.5 | 991  | 2 Q6NZM2     | Q6NZM2 mus musculu |
| 43 | 763   | 4.5 | 991  | 2 AAH66062   | AAH66062 mus muscu |
| 44 | 762.5 | 4.5 | 1007 | 2 Q8J128     | Q8J128 xenopus lae |
| 45 | 761   | 4.5 | 1024 | 1 SZ6L_HUMAN | Q9b9h1 homo sapien |

#### ALIGNMENTS

|          |  |          |               |
|----------|--|----------|---------------|
| RESULT 1 | CSM2_HUMAN   | STANDARD | PRT, 3487 AA. |
| ID       | Q72408; Q8N963; Q96C03; Q9H4V7; Q9H4V8; Q9H4V9; Q9H4W0; Q9H4W1;        |          |               |
| AC       | Q9H4W2; Q9H4W3; Q9H4W4; Q9HCY5; Q9HCY6; Q9HCY7;                        |          |               |
| DT       | 29-MAR-2004 (Ref. 43, Created)   |          |               |
| DT       | 29-MAR-2004 (Ref. 43, Last sequence update)                            |          |               |
| DT       | 05-JUL-2004 (Ref. 44, Last annotation update)                          |          |               |
| DE       | CUB and sushi multiple domains protein 2.                              |          |               |
| GN       | Name=CSMD2; Synonyms=KIAA1884;   |          |               |
| OS       | Homo sapiens (Human).  |          |               |
| OC       | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;      |          |               |
| OC       | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.              |          |               |
| OX       | NCBI_TaxID=9606;   |          |               |
| XP       | [1]  |          |               |
| RP       | SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.                |          |               |
| RX       | MEHLIE=22788796; PubMed=12906867; DOI=10.1016/S0888-7543(03)00149-6;   |          |               |
| RA       | Lau W.L., Scholnick S.B.;  |          |               |
| RT       | "Identification of two new members of the CSMD gene family.";          |          |               |
| RL       | Genomics 82:412-415(2003).   |          |               |
| RN       | [2]  |          |               |
| RP       | SEQUENCE FROM N.A. (ISOFORM 2), AND SEQUENCE OF 2969-3487 (ISOFORM 1). |          |               |
| RC       | TISSUE=Brain, and Testis; carcinoma;                                   |          |               |
| RX       | PubMed=14702039; DOI=10.1038/ng1285;                                   |          |               |
| RA       | Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,      |          |               |
| RA       | Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,     |          |               |
| RA       | Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,    |          |               |
| RA       | Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,             |          |               |
| RA       | Nagahori K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,       |          |               |
| RA       | Shiratori A., Sudo H., Hosogi T., Kaku Y., Kodaira H., Kondo H.,       |          |               |
| RA       | Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  |          |               |
| RA       | Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,       |          |               |
| RA       | Yamazaki M., Nishimura K., Ishibashi T., Yamashita H., Murakawa K.,    |          |               |
| RA       | Fujimori K., Tanai H., Kimata M., Matsumura M., Hiraoka S., Chiba Y.,  |          |               |
| RA       | Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Horiuchi T., |          |               |
| RA       | Kusano J., Kanehori K., Takahashi-Fujii A., Hara K., Tanabe T.-O.,     |          |               |
| RA       | Imura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,         |          |               |
| RA       | Imose N., Mutsaers K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,      |          |               |
| RA       | Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,         |          |               |
| RA       | Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  |          |               |
| RA       | Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,   |          |               |
| RA       | Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,      |          |               |
| RA       | Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,         |          |               |
| RA       | Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiiwa T.,         |          |               |
| RA       | Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,             |          |               |
| RA       | Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., |          |               |
| RA       | Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Sena T.,      |          |               |
| RA       | Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,          |          |               |
| RA       | Togashi T., Oyama J., Hata H., Watanabe M., Komatsu T., Nakagawa K.,   |          |               |
| RA       | Mitsushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Yamashita R., |          |               |
| RA       | Okumura K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamashita R.,  |          |               |
| RA       | Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;        |          |               |
| RT       | "Complete sequencing and characterization of 21,243 full-length human  |          |               |
| CDNAs."; |  |          |               |

RL Nat. Genet. 36:40-45(2004).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RA Wallis J., Brown A.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RC TISSUE=Brain;  
 RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stalcup M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci F., Prange C.,  
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Hellon E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywicki M.I., Skalska U., Smallus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RP (5)  
 RN SEQUENCE OF 2542-3487 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.  
 RC TISSUE=Brain;  
 RX MEDLINE=21456161; PubMed=11572484;  
 RA Nagase T., Kikuno R., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XXI.  
 RT The complete sequences of 60 new cDNA clones from brain which code for  
 RT large proteins.";  
 RL DNA Res. 8:179-187(2001).  
 RL -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=Q72408-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q72408-2; Sequence=VSP\_009038, VSP\_009039, VSP\_009040,  
 CC VSP\_009041, VSP\_009042;  
 CC Note=No experimental confirmation available;  
 CC Name=3;  
 CC IsoId=Q72408-3; Sequence=VSP\_009043, VSP\_009044, VSP\_009045,  
 CC VSP\_009046;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Weakly expressed in most tissues, except in  
 CC brain. Expressed at intermediate level in brain, including  
 CC cerebellum, substantia nigra, hippocampus and fetal brain.  
 CC Overexpressed in some head and neck cancer cell lines.  
 CC -1- SIMILARITY: Belongs to the CSMD family.  
 CC -1- SIMILARITY: Contains 14 CUB domains.  
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to a  
 CC frameshift in position 939 that shortens the protein by 1021  
 CC residues in its N-terminus. It is unknown whether the sequence  
 CC shown exists or whether Ref.3 is right, shortening the sequence in  
 CC its N-terminus.  
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AY10418; AAC34701.1; -.  
 DR EMBL; AK095627; BAC04593.1; ALT\_INIT.  
 DR EMBL; AK127722; BAC87101.1; -.  
 DR EMBL; AC115285; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AC115286; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AL121980; CAC05319.1; ALT\_SEQ.  
 DR EMBL; AL121980; CAC05320.1; ALT\_SEQ.  
 DR EMBL; AL121980; CAC05321.1; ALT\_SEQ.  
 DR EMBL; AL121980; CAC05322.1; ALT\_SEQ.  
 DR EMBL; AL355178; CAC10283.1; ALT\_SEQ.  
 DR EMBL; AL355178; CAC10284.1; ALT\_SEQ.  
 DR EMBL; AL355178; CAC10285.1; ALT\_SEQ.  
 DR EMBL; AL355178; CAC10286.1; ALT\_SEQ.  
 DR EMBL; AL355178; CAC10287.1; ALT\_SEQ.  
 DR EMBL; AL355178; CAC10288.1; ALT\_SEQ.  
 DR EMBL; AL355178; CAC10289.1; ALT\_SEQ.  
 DR EMBL; AL355178; CAC10290.1; ALT\_SEQ.  
 DR EMBL; AL596824; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AL607106; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; BC031871; AH31871.1; -.  
 DR EMBL; AB067471; BAB67777.1; -.  
 DR HSSP; P10998; 1VD.  
 DR Genew; HGNC:19290; CSMD2.  
 DR MIM; 608398; -.  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF00431; CUB; 14.  
 DR Pfam; PF00084; Sushi; 26.  
 DR PROSITE; PS01180; CUB; 14.  
 DR PROSITE; PS0923; SUSHI; 26.  
 KW Alternative splicing; Direct protein sequencing; Repeat; Sushi;  
 KW Transmembrane.  
 FT DOMAIN 1 3408 Extracellular (Potential).  
 FT TRANSMEM 3409 3429 Potential.  
 FT DOMAIN 3430 3487 Cytoplasmic (Potential).  
 FT DOMAIN 26 134 CUB 1.  
 FT DOMAIN 137 198 Sushi 1.  
 FT DOMAIN 202 306 Sushi 2.  
 FT DOMAIN 341 402 Sushi 2.  
 FT DOMAIN 405 516 Sushi 3.  
 FT DOMAIN 519 576 Sushi 3.  
 FT DOMAIN 578 686 Sushi 4.  
 FT DOMAIN 689 750 Sushi 4.  
 FT DOMAIN 752 860 Sushi 5.  
 FT DOMAIN 865 922 Sushi 5.  
 FT DOMAIN 924 1034 CUB 6.  
 FT DOMAIN 1037 1096 Sushi 6.  
 FT DOMAIN 1098 1206 CUB 7.  
 FT DOMAIN 1209 1269 Sushi 7.  
 FT DOMAIN 1271 1380 CUB 8.  
 FT DOMAIN 1383 1443 Sushi 8.  
 FT DOMAIN 1445 1553 CUB 9.  
 FT DOMAIN 1556 1617 Sushi 9.  
 FT DOMAIN 1619 1727 CUB 10.  
 FT DOMAIN 1733 1794 Sushi 10.  
 FT DOMAIN 1796 1904 CUB 11.  
 FT DOMAIN 1907 1966 Sushi 11.  
 FT DOMAIN 1968 2076 CUB 12.  
 FT DOMAIN 2121 2180 Sushi 12.  
 FT DOMAIN 2182 2293 CUB 13.  
 FT DOMAIN 2292 2353 Sushi 13.  
 FT DOMAIN 2355 2466 CUB 14.  
 FT DOMAIN 2466 2528 Sushi 14.  
 FT DOMAIN 2529 2590 Sushi 15.  
 FT DOMAIN 2591 2655 Sushi 15.  
 FT DOMAIN 2656 2713 Sushi 16.  
 FT DOMAIN 2714 2771 Sushi 17.  
 FT DOMAIN 2772 2829 Sushi 18.  
 FT DOMAIN 2833 2890 Sushi 19.  
 FT DOMAIN 2891 2949 Sushi 20.  
 FT DOMAIN 2950 3009 Sushi 21.  
 FT DOMAIN 3010 3067 Sushi 22.  
 FT DOMAIN 3068 3125 Sushi 23.  
 FT DOMAIN 3125 Sushi 24.

|  |        |                        |   |           |  |
|--|--------|------------------------|---|-----------|--|
| FT   | DOMAIN | 3129                   | 3187  | Suehi 25. |  |
| FT   | DOMAIN | 3188                   | 3247  | Suehi 26. |  |
| Query Match  |        |                        |   |           | 90.5%; Score 15364.5; DB 1; Length 3487; |
| Best Local Similarity  |        |                        |   |           | 91.9%; Pred. No. 0;                      |
| Matches 2847; Conservative 12; Mismatches 39; Indels 199; Gaps 10; |        |                        |   |           |  |
| Qy   | 2      | AGAPPPL---             | LLPCSLISDCCASNORHSVGPSELVYKQIELEKSGVLTMPKNSQ    | 57        |  |
| Db   | 271    | ASLPAPVYSSKMWLRHFTSD-- | GNHRQR-GFSAQYOVKKQIELEKSGVLTMPKNSQ              | 326       |  |
| Qy   | 58     | KTSVLTQGVSGQHNMCPDPCI  | PERGKRLASDFRLGSSVQFTCEGYDQSGKEITCKV             | 117       |  |
| Db   | 327    | KTSVLTQGVSGQHNMCPDPCI  | PERGKRLASDFRLGSSVQFTCEGYDQSGKEITCKV             | 386       |  |
| Qy   | 118    | SDMPAASDHRPVCARMCDHLR  | PGSIIITSPNFIQYDNNAHCIWITLANPSKYIKL              | 177       |  |
| Db   | 387    | SDMPAASDHRPVCARMCDHLR  | PGSIIITSPNFIQYDNNAHCIWITLANPSKYIKL              | 446       |  |
| Qy   | 178    | AFFEPDLRGYDTLVGDGDGDK  | KTLYVMSQACSDSPHTPSRIIPESMSDIPQKX                | 237       |  |
| Db   | 447    | AFFEPDLRGYDTLVGDGDGDK  | KTLYVMSQACSDSPHTPSRIIPESMSDIPQKX                | 495       |  |
| Qy   | 238    | TVLEICDISSDARSQSVRS    | PKTSNAVELVAPGTEIEQSCDPCIIPAYGREGSRFH            | 297       |  |
| Db   | 496    | LLFG-----TDG-SGS-----  | SLGFKASVEIEEQSCDPCIIPAYGREGSRFH                 | 538       |  |
| Qy   | 298    | HGDTLKECOPAFELVGOKAIT  | COKNQWSAKKQCVSCFPNFTSPSGVLSNYPEDY               | 357       |  |
| Db   | 539    | HGDTLKECOPAFELVGOKAIT  | COKNQWSAKKQCVSCFPNFTSPSGVLSNYPEDY               | 598       |  |
| Qy   | 358    | GNHLHCWLLIARPSRIHLA    | FNIDIVRPODFLVIKQATAEAPVLTGFGNDLPSSIT            | 417       |  |
| Db   | 599    | GNHLHCWLLIARPSRIHLA    | FNIDIVRPODFLVIKQATAEAPVLTGFGNDLPSSIT            | 658       |  |
| Qy   | 418    | SSGHVAREFQTDHSTGKRG    | FNITFTFRHNECPDPGVNPKRGFSDLSQSSISFLCD            | 477       |  |
| Db   | 659    | SSGHVAREFQTDHSTGKRG    | FNITFTFRHNECPDPGVNPKRGFSDLSQSSISFLCD            | 718       |  |
| Qy   | 478    | EGFLTGQSEITITCVLKEG    | SVVMSAVLRCEAPCGHLTSPSGTILSPGMPGYKDALSC          | 537       |  |
| Db   | 719    | EGFLTGQSEITITCVLKEG    | SVVMSAVLRCEAPCGHLTSPSGTILSPGMPGYKDALSC          | 778       |  |
| Qy   | 538    | AMVLEAOGYPIKITTDFR     | KTEVNYDTLEVNDGRYSAPLIGVHGTQVPOFLSTSYL           | 838       |  |
| Db   | 779    | AMVLEAOGYPIKITTDFR     | KTEVNYDTLEVNDGRYSAPLIGVHGTQVPOFLSTSYL           | 838       |  |
| Qy   | 598    | YLLFSTDKSHDIDIGFOL     | RYETITLOSCHLDPGIPVNGQRHGNDFYVGAULTFSCDSGYT      | 657       |  |
| Db   | 839    | YLLFSTDKSHDIDIGFOL     | RYETITLOSCHLDPGIPVNGQRHGNDFYVGAULTFSCDSGYT      | 898       |  |
| Qy   | 658    | LSDEPLECEBPNFQWSRAL    | PSCEALCGGFIQSSSGTILSPGPDPFPYNNLCTWIIETSH        | 717       |  |
| Db   | 899    | LSDEPLECEBPNFQWSRAL    | PSCEALCGGFIQSSSGTILSPGPDPFPYNNLCTWIIETSH        | 958       |  |
| Qy   | 718    | GKGVFTFTHTHLSSGHY      | LLITENGSTFQRLQLTGSRLPAPISAGIYGNTQVRRIS          | 777       |  |
| Db   | 959    | GKGVFTFTHTHLSSGHY      | LLITENGSTFQRLQLTGSRLPAPISAGIYGNTQVRRIS          | 1018      |  |
| Qy   | 778    | DFSNSYEGFNITFSBYD      | LEPCBEPVPAYSIRKGLQFVGDTLTFSCPGRLSGTARIT         | 837       |  |
| Db   | 1019   | DFSNSYEGFNITFSBYD      | LEPCBEPVPAYSIRKGLQFVGDTLTFSCPGRLSGTARIT         | 1078      |  |
| Qy   | 838    | CLGGRRLWSSPLPRCA       | ECNSVYTGOTLLSPNFPVNNNNHCCIYSIGTOPGKGIOL         | 897       |  |
| Db   | 1079   | CLGGRRLWSSPLPRCA       | ECNSVYTGOTLLSPNFPVNNNNHCCIYSIGTOPGKGIOL         | 1138      |  |
| Qy   | 898    | KARAFELSEGDUVKYD       | NNNSARLLGVFSHSEMMGTLLNSTSSLMWDTITDAENSKG        | 957       |  |
| Db   | 1139   | KARAFELSEGDUVKYD       | NNNSARLLGVFSHSEMMGTLLNSTSSLMWDTITDAENSKG        | 1198      |  |
| Qy   | 958    | FELHFSFELIKCEDP        | TPKFGYKVHDBGHFAGSSVSFCDPGYSLRGEELLCLSGERR       | 1017      |  |
| Db   | 1199   | FELHFSFELIKCEDP        | TPKFGYKVHDBGHFAGSSVSFCDPGYSLRGEELLCLSGERR       | 1258      |  |
| Qy   | 1018   | TWBRPLPTCAECGGT        | TRGEVSGQVLSPGYAPVREHNLNCIWTTEARAGCTIGHPLVFD     | 1077      |  |
| Db   | 1259   | TWBRPLPTCAECGGT        | TRGEVSGQVLSPGYAPVREHNLNCIWTTEARAGCTIGHPLVFD     | 1318      |  |
| Qy   | 1078   | TEEVHDLRIWIDG          | VEGSGVLLKELSGPALPKDLSTNSVVLQSTDFPSTSOGFAIOFS    | 1137      |  |
| Db   | 1319   | TEEVHDLRIWIDG          | VEGSGVLLKELSGPALPKDLSTNSVVLQSTDFPSTSOGFAIOFS    | 1378      |  |
| Qy   | 1138   | VSTATSCNDPCI           | PONGSRSDSWAGDSTVPOCPGVALQCSAEISCVKIEENRFWQSP    | 1197      |  |
| Db   | 1379   | VSTATSCNDPCI           | PONGSRSDSWAGDSTVPOCPGVALQCSAEISCVKIEENRFWQSP    | 1438      |  |
| Qy   | 1198   | PTCIAPCGDLTGSSG        | VITLSPNPEPPEPKCECMKTVSPDYIALVFNIFNLPEGYDF       | 1257      |  |
| Db   | 1439   | PTCIAPCGDLTGSSG        | VITLSPNPEPPEPKCECMKTVSPDYIALVFNIFNLPEGYDF       | 1498      |  |
| Qy   | 1258   | LHIYDGRDLSPLIGS        | FGSOLPGRIESSNSLFLAFRSDASVSNAGFVIDYTENPRESC      | 1317      |  |
| Db   | 1499   | LHIYDGRDLSPLIGS        | FGSOLPGRIESSNSLFLAFRSDASVSNAGFVIDYTENPRESC      | 1558      |  |
| Qy   | 1318   | FDPGSIKNGTRVGS         | DLKLGSVITYYCHGYEVEGTSTLSCILGPDGKPVNNRPVCTAP     | 1377      |  |
| Db   | 1559   | FDPGSIKNGTRVGS         | DLKLGSVITYYCHGYEVEGTSTLSCILGPDGKPVNNRPVCTAP     | 1618      |  |
| Qy   | 1378   | CGGQYVSSDGVVLS         | PNTPONTSGOICLYPYTVPKOVVVGQAFEFHTALNDVYVHDG      | 1437      |  |
| Db   | 1619   | CGGQYVSSDGVVLS         | PNTPONTSGOICLYPYTVPKOVVVGQAFEFHTALNDVYVHDG      | 1678      |  |
| Qy   | 1438   | SOHSRLSSLSGSH          | TGSLPLATSNQVLIKPSAKGLAPARGFHVQYAVPRTSAQCSSV     | 1497      |  |
| Db   | 1679   | SOHSRLSSLSGSH          | TGSLPLATSNQVLIKPSAKGLAPARGFHVQYAVPRTSAQCSSV     | 1738      |  |
| Qy   | 1498   | PEPRYGRRLGSD           | SEFVGAIVAFECNSGYALQCSPEIECLPVGALAQNVSAPTCVPCGG  | 1557      |  |
| Db   | 1739   | PEPRYGRRLGSD           | SEFVGAIVAFECNSGYALQCSPEIECLPVGALAQNVSAPTCVPCGG  | 1798      |  |
| Qy   | 1558   | NLTERRGITLSPG          | PEPYLANSKNCWKIIVPEGAGIOIOVVSFTTEBQNMBSLRFPGDADN | 1617      |  |
| Db   | 1799   | NLTERRGITLSPG          | PEPYLANSKNCWKIIVPEGAGIOIOVVSFTTEBQNMBSLRFPGDADN | 1858      |  |
| Qy   | 1618   | TVTMLGFSFGT            | TVPALNSTSNOLYLHFYSDISVSAAGFHELYTVGSSCPPEAVPNSG  | 1677      |  |
| Db   | 1859   | TVTMLGFSFGT            | TVPALNSTSNOLYLHFYSDISVSAAGFHELYTVGSSCPPEAVPNSG  | 1918      |  |
| Qy   | 1678   | VKTGERYLVNDVVS         | FOCEPGYALQGHAIISCMPTVARMMYPPPLCIAOCGGIVEEMGV    | 1737      |  |
| Db   | 1919   | VKTGERYLVNDVVS         | FOCEPGYALQGHAIISCMPTVARMMYPPPLCIAOCGGIVEEMGV    | 1978      |  |
| Qy   | 1738   | ILSPGPGNT              | PSNMDGSKIALPVFGAHIOFLNFSTEPNHDIYEINRGPYETSRMGRF | 1797      |  |
| Db   | 1979   | ILSPGPGNT              | PSNMDGSKIALPVFGAHIOFLNFSTEPNHDIYEINRGPYETSRMGRF | 2038      |  |
| Qy   | 1798   | SGSELPSLSLST           | SHETTVYFHSDBSONRPGFKLEYO-----                   | 1833      |  |
| Db   | 2039   | SGSELPSLSLST           | SHETTVYFHSDBSONRPGFKLEYO-----                   | 2098      |  |
| Qy   | 1834   | -----AYELOCPD          | PEPFANGIVGAGYNGQSYTFBCLPCYQULGH                 | 1875      |  |
| Db   | 2099   | STPPVAASVYWD           | LDPGCAVELOCPDPEPPRANGIVAGAGYNGQSYTFBCLPCYQULGH  | 2158      |  |
| Qy   | 1876   | PVLTQGHTRNMDH          | PLPKCEVPCGNGITSSNGTVYSPGPPSPYSSQDCWMLITVPIGH    | 1935      |  |
| Db   | 2159   | PVLTQGHTRNMDH          | PLPKCEVPCGNGITSSNGTVYSPGPPSPYSSQDCWMLITVPIGH    | 2218      |  |
| Qy   | 1936   | GVRNLNLSLLOTE          | PSGDIITLWIDGPOOTAPRLGVTRSMARKTVYSSNOVLLKXHRDAAT | 1995      |  |
| Db   | 2219   | GVRNLNLSLLOTE          | PSGDIITLWIDGPOOTAPRLGVTRSMARKTVYSSNOVLLKXHRDAAT | 2278      |  |
| Qy   | 1996   | GGIPLAIFSAVPLTK        | CPPTILPNAEVVTEENENIDIVRYRCLPGLFTLVGNEILLTKL     | 2055      |  |
| Db   | 2279   | GGIPLAIFSAVPLTK        | CPPTILPNAEVVTEENENIDIVRYRCLPGLFTLVGNEILLTKL     | 2338      |  |

QY 2056 GTVLOEGEPPICEVHCPTNELLTDS TGVILSOSVPGSVPOFOTCSMLVVRPVDNIELT 2115  
 DB 2339 GTVLOEGEPPICEVHCPTNELLTDS TGVILSOSVPGSVPOFOTCSMLVVRPVDNIELT 2398  
 QY 2116 VEYFISEKQYDEFEIFDGSQSPILKALSGNYSAPLIVTSSSNVLRMSDHAHNKG 2175  
 DB 2399 VEYFISEKQYDEFEIFDGSQSPILKALSGNYSAPLIVTSSSNVLRMSDHAHNKG 2458  
 QY 2176 FIKRSAPYCSLPRLPIHGLITGOTSTQSGSIHGCNAGRVLVHSMALCTRRHPQYHL 2235  
 DB 2459 FIKRSAPYCSLPRLPIHGLITGOTSTQSGSIHGCNAGRVLVHSMALCTRRHPQYHL 2518  
 QY 2236 WSEAIPLCOLALSCGLPEAPKNGMVEGKEYTVGTRKAVSCSEGYHLQAGABATAECLDTGL 2295  
 DB 2519 WSEAIPLCOLALSCGLPEAPKNGMVEGKEYTVGTRKAVSCSEGYHLQAGABATAECLDTGL 2578  
 QY 2236 WSNRVVPQCVVTPDVVSSISVEHGRNRLIFETQYFOAOLMLICDGYYYTGORVIRC 2355  
 DB 2579 WSNRVVPQCVVTPDVVSSISVEHGRNRLIFETQYFOAOLMLICDGYYYTGORVIRC 2638  
 QY 2356 QANGKMSIGDSTPTCRITISGCELPPIPNGHRIIGLISVYGATATSCNSGYTLVSSRREC 2415  
 DB 2639 QANGKMSIGDSTPTCRITISGCELPPIPNGHRIIGLISVYGATATSCNSGYTLVSSRREC 2698  
 QY 2416 MANGIWSGEVRCIAGHCCTPEPIVNGHINENYSYSGSVVYQCNAGFRILGMSVRICQ 2475  
 DB 2699 MANGIWSGEVRCIAGHCCTPEPIVNGHINENYSYSGSVVYQCNAGFRILGMSVRICQ 2758  
 QY 2476 DHMSGKPCFCVPIITCGHRGNPANGLOGNOPINADVKEPCNPGYNAEGANSOCLASG 2535  
 DB 2759 DHMSGKPCFCVPIITCGHRGNPANGLOGNOPINADVKEPCNPGYNAEGANSOCLASG 2770  
 QY 2536 QMSDMLPTCRITINCTDPCGHQENSVROVHAGPHRFSGTIVSYNCHGFYILGTPTVLSCQ 2595  
 DB 2771 QMSDMLPTCRITINCTDPCGHQENSVROVHAGPHRFSGTIVSYNCHGFYILGTPTVLSCQ 2770  
 QY 2596 GDGTWDRPRPQCLLVSCGHPSSPHSQSGSYTVGAVVAVYSCIGKRTLVGNSTRMCGLD 2655  
 DB 2771 GDGTWDRPRPQCLLVSCGHPSSPHSQSGSYTVGAVVAVYSCIGKRTLVGNSTRMCGLD 2817  
 QY 2656 GHMTGSLPHCSGTSVYCGADPGIPAHGIRIGDSFDPCTWRFSEAGHVTLRGSSERTCA 2715  
 DB 2818 GHMTGSLPHCSGTSVYCGADPGIPAHGIRIGDSFDPCTWRFSEAGHVTLRGSSERTCA 2877  
 QY 2716 NGWSGSGQPECGVISCNPNPTSPNARVYFSDGLVFSSSIYEGREGYATGLSRCSVA 2775  
 DB 2878 NGWSGSGQPECGVISCNPNPTSPNARVYFSDGLVFSSSIYEGREGYATGLSRCSVA 2937  
 QY 2776 GTWGSDEPCIVNCGDPI PANGLRLGNDPRYNTKTYTTCVGVYMMESHRSVLSCTD 2835  
 DB 2938 GTWGSDEPCIVNCGDPI PANGLRLGNDPRYNTKTYTTCVGVYMMESHRSVLSCTD 2997  
 QY 2836 RYNGTKPVCKALMKCPPLIPNGKVVSGDFMWSGSSVYACLEGYLSLPAVTCGNGS 2895  
 DB 2998 RYNGTKPVCKALMKCPPLIPNGKVVSGDFMWSGSSVYACLEGYLSLPAVTCGNGS 3057  
 QY 2896 WTEBELPQCFVFCGDRVPSRGRREDRGSYSSVSFSCHPPLIVLGSRRRRCQSDGWS 2955  
 DB 3058 WTEBELPQCFVFCGDRVPSRGRREDRGSYSSVSFSCHPPLIVLGSRRRRCQSDGWS 3117  
 QY 2956 GTOPSCIDPFLTTCADBGVPOFGIIONNSOGYGVSTVLRCKCKGYTLQOSTTRTCLPNT 3015  
 DB 3118 GTOPSCIDPFLTTCADBGVPOFGIIONNSOGYGVSTVLRCKCKGYTLQOSTTRTCLPNT 3177  
 QY 3016 WSGTPDPCVPHHCRQETPTTHANVGLDPSMGYTLI 3052  
 DB 3178 WSGTPDPCVPHHCRQETPTTHANVGLDPSMGYTLI 3214

RESULT 2  
 CSM1\_HUMAN  
 ID CSM1\_HUMAN STANDARD; PRT; 3565 AA.

AC Q96P27; Q96CU9; Q96RM4;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE CUB and sushi multiple domains protein 1 precursor (U05952/PRO19863).  
 GN Name-CSM1; Synonyms=K1AA1890;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=21365705; PubMed=1172063; DOI=10.1006/geno.2001.6587;  
 RA Sun P.C., Uppaluri R., Schmidt A.P., Pashia M.E., Quant E.C.,  
 RA Sunwoo J.B., Gollin S.M., Scholnick S.B.;  
 RT "transcript map of the 8p23 putative tumor suppressor region.";  
 RL Genomics 75:17-25(2001).  
 RN [2]  
 RP SEQUENCE OF 1319-3565 FROM N.A. (ISOFORM 3).  
 RC TISSUE=Brain;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Oca T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
 RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
 RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,  
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,  
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Iosida M., Hota T.,  
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
 RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,  
 RA Imose N., Musashino K., Yuki H., Oshima A., Sasaki N., Aotsuma S.,  
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
 RA Moriya S., Momiyama H., Satoh N., Takami S., Teraishi Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Ohtani K., Wakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togoashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 RN [3]  
 RP SEQUENCE OF 825-3565 FROM N.A. (ISOFORM 4).  
 RC MEDLINE=22887296; PubMed=12975309; DOI=10.1101/gr.1293003;  
 RX Clead H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,  
 RA Chen J., Chow B., Chui C., Crowley C., Currell B., Dowd P.,  
 RA Eaton D., Foster J., Grimaldi C., Gu Q., Hass P.E., Heldens S.,  
 RA Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,  
 RA Lewis L., Liao D., Mark M., Kobbie B., Sanchez C., Schoenfeld J.,  
 RA Seisagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagtes A.,  
 RA Vanden R., Watanabe C., Wleand D., Woods K., Xie M.-H., Yansura D.,  
 RA Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,  
 RA Godowski P., Gray A.;  
 RT "The secreted protein discovery initiative (SPDI), a large-scale  
 RT effort to identify novel human secreted and transmembrane proteins: a  
 RT bioinformatics assessment.";  
 RL Genome Res. 13:2265-2270(2003).  
 RN [4]  
 RP SEQUENCE OF 966-2013 FROM N.A. (ISOFORM 4), AND TISSUE SPECIFICITY.  
 RC TISSUE=Brain;  
 RX MEDLINE=21456161; PubMed=11572484;  
 RA Nagase T., Kikuno R., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XXI.  
RT The complete sequences of 60 new cDNA clones from brain which code for  
RT large proteins.";  
RL DNA Res. 8:179-187(2001).  
RN [5]  
RP DISEASE.  
RX PubMed=12696061; DOI=10.1002/gcc.10191;  
RA Toomes C., Jackson A., Maguire K., Wood J., Gollin S., Ishwad C.,  
RA Peterson I., Prime S., Parkinson K., Bell S., Woods G., Matham A.,  
RA Oliver R., Woodward R., Sloan P., Dixon M., Read A., Thakker N.,  
RT "The presence of multiple regions of homozygous deletion at the CSM1  
RT locus in oral squamous cell carcinoma question the role of CSM1 in  
RT head and neck carcinogenesis.";  
RL Genes Chromosomes Cancer 37:132-140(2003).  
RP [6]  
RX DISEASE.  
RX PubMed=14506705; DOI=10.1002/gcc.10279;  
RA Scholnick S.B., Richter T.M.,  
RT "The role of CSM1 in head and neck carcinogenesis.";  
RL Genes Chromosomes Cancer 38:281-283(2003).  
CC -1- FUNCTION: Potential suppressor of squamous cell carcinomas.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Name=1;  
CC IsoId=Q96P27-1; Sequence=Displayed;  
CC Name=2; Synonym=Short;  
CC IsoId=Q96P27-2; Sequence=VSP\_009034, VSP\_009035;  
CC Name=3;  
CC IsoId=Q96P27-3; Sequence=VSP\_009030, VSP\_009031;  
CC Note=No experimental confirmation available;  
CC Name=4;  
CC IsoId=Q96P27-4; Sequence=VSP\_009032, VSP\_009033;  
CC -1- TISSUE SPECIFICITY: Weakly expressed in most tissues, except in  
CC brain. Expressed at intermediate level in brain, including  
CC cerebellum, substantia nigra, hippocampus and fetal brain.  
CC DISEASE: Defects in CSM1 may be a cause of oral and oropharyngeal  
CC squamous cell carcinomas (OSCCs). Ref.5 and Ref.6 are however in  
CC disagreement: while Ref.6 considers CSM1 as a strong candidate  
CC for OSCCs, Ref.5 thinks it is not.  
CC -1- SIMILARITY: Belongs to the CSM1 family.  
CC -1- SIMILARITY: Contains 14 CUB domain.  
CC -1- SIMILARITY: Contains 28 Sushi (CCP/SCR) domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL; AF333704; AAK73475.2; -;  
DR EMBL; AY017307; AAG52948.1; -;  
DR EMBL; AK126936; BAC86754.1; ALT\_INIT.  
DR EMBL; AY358174; AA088541.1; ALT\_INIT.  
DR EMBL; AB067477; BAB67783.1; -;  
DR Genew; HGNC:14026; CSM1.  
DR MIM; 608397; -;  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
DR Pfam; PR000431; CUB; 6.  
DR Pfam; PF00084; Sushi; 6.  
DR PROSITE; PS0180; CUB; 14.  
DR PROSITE; PS50923; SUSHI; 28.  
KM Alternative splicing; Repeat; Signal; Sushi; Transmembrane.  
FT SIGNAL 1 26 Potential.  
FT CHAIN 27 3565 CUB and sushi multiple domains protein 1.  
FT DOMAIN 27 3488 Extracellular (Potential).  
FT TRANSMEM 3489 3509 Cytoplasmic (Potential).  
FT DOMAIN 3510 3565 Potential.  
FT DOMAIN 32 140 CUB 1.  
FT DOMAIN 143 204 Sushi 1.

| Qy | Db | 1   | MAGAPPALP-----LPCSLISD-----CCASQNRHVGVPBELYKQTELKSRGV        | 47  |
|----|----|-----|--|-----|
| Qy | Db | 264 | ISCTEAPSIWLTGMLNLPSPVITSSKNWLRHLFTSDSHRRK-GNAGQFQVKAIELKSRGV | 322 |
| Qy | Db | 48  | KLMPKDNQKTVLVTVQVGSQHNNCPDPRGIERGRKLSDP-RIGSSVQFTCNBGYD      | 106 |
| Qy | Db | 323 | KMLPSKXGSHKNSVLSQGVAVLSHMCPLPGIPIENGRRRAGSSDFSRGANVQFSCBNVYL | 382 |
| Qy | Db | 107 | QGSKRITCKVMSFPAWSDHPRVPCARBMCDALRGPSGIIITSPNPPIQVNNAHCVWII   | 166 |
| Qy | Db | 383 | QGSKSIITCORVITTLTAMSDHPIICARITGSLKRGSGVITSPNPPIQVNNAHCVWII   | 442 |
| Qy | Db | 167 | TALNPSKVIKLAPEERDLERGYDTLVGDDGQDGDQKTVLYMSQNAQCSDPHTPGRIPE   | 226 |
| Qy | Db | 443 | TTTDPDKVITLAPEERDLERGYDTLVGDDGQDGDQKTVLYV-----LTGSSVDP       | 492 |

Query Match 67.0%; Score 11372.5; DB 1; Length 3565;  
Best Local Similarity 64.1%; Pred. No. 0;  
Matches 1970; Conservative 461; Mismatches 584; Indels 59; Gaps 9;

QY 227 ---SMSGDIWROKWTVLEICNDISSDARSQSVKSPKTSNAVELVAG-----TEHEOG 278  
 DB 493 LIVSNQNMWLMH-----LQSDG-----IGSPGFAVYVQZIEKG 526  
 QY 279 SCGDGIPAYGRREGSRPHGDDTLKECOPAFELVQKAITCOKNNQSAKPGCVSPSC 338  
 DB 527 GCGDGIIPAYGRKTSSTFLHGDTLFECRPAFELVGERVITCQNNQMSGKRECVSCF 586  
 QY 339 FNFSPGCVLSPNVPEDYGNHLHCWMLILARPESRHLAANDIDVEPODFELVIXGAT 398  
 DB 567 FNFSTSGSILISPNPEEYGNMNCWMLIIEPGRILHLPNDPVEPODFELVIXGDI 646  
 QY 399 AEAPLYGTFSGNQLPSSITSSGHVAREFOTDHSNGKGFNITETTERHNCBPGRVYN 458  
 DB 647 SDITVLTGTFSGNEVPSCASSGHVLEFQSDHSTTGGRFNITYITTGQNECHDPGIPIN 706  
 QY 459 GKRFDLSIQGSSISFLCDEGFLTGSGEITTCVLKGSVVMNSAVLRCEAPCGHLLTSP 518  
 DB 707 GRRGDRFLGSSVSFHCDDGFVKTQSGESITCLOGNVWMSITVRCBAPCGHLLTAS 766  
 QY 519 SGTILSPGMPGFYDALSACAWIEAOGCPYKITEFDPKTEVNYDTLEFVDRGRTYSAPLI 578  
 DB 767 SGVILPPEMPGPIYKOSLHCWMIIEAKPGHSTIKMTFDRFOTEVNYDTLEVRDGPASSSPLI 826  
 QY 579 GVHGTQVPOPLISTSNVLYLFFSTDKSHSDIGFQLREYETTLQSDHCLDPGI PVNGQRH 638  
 DB 827 GEYHGTQAPQLISTGNFMWLLFTTDSRSISIGLHYESVTLESDBCLDPGIPVNGHRH 886  
 QY 639 GNDYVVALYTFSCDSGYTLSDGEPLCEBNFQMSRALPSCCALCGGFIQSGSGTILISPG 698  
 DB 887 GGDGIGISTVTFSCDPPYTLSDDEPLVCERNHQNHALPSCDALCGGFIQSGSGTILISPG 946  
 QY 699 FPDPEYPNNLNCTWIIETSHGKGVFTFHTPHLSGHDYLLITNGSGTOQLRGLTSSRLP 758  
 DB 947 FPDPEYPNNLNCTWIIETSHGKGVFTFHTPHLSGHDYLLITNGSGTOQLRGLTSSRLP 758  
 QY 759 APISAGLYGNPFAOVRFISDFSMSYEGFNITFSEYDLEPCERPEVPAVYSIRKGLQGVAD 818  
 DB 1007 HTIKAGJFGNFTAOIRISDFSISYEGFNITFSEYDLEPCERPEVPAVYSIRKGLQGVAD 1066  
 QY 819 TLFPSCPGVRLGSLTAITLGGRRRLMSPLRCAVACNSITGQGTILISNPNFVYN 878  
 DB 1067 SLTFSCGLYRLGATYLTCLGGRRRWSAPLPRCAVACNSITGQGTILISNPNFVYN 1126  
 QY 879 NNHECIYSIQTOPGKGIQAKARAFELSEGDLKVYDGNNSARLLGVFSHEMKGVTILNS 938  
 DB 1127 NNHECIYSIQTOPGKGIQAKARAFELSEGDLKVYDGNNSARLLGVFSHEMKGVTILNS 938  
 QY 939 TSSSLWDLFTTDAENTSKGFEHLFSSPELICEDEPCTPKRGYKVHDEGHFAGSSVSFSCD 998  
 DB 1187 TSSSLWDLFTTDAENTSKGFEHLFSSPELICEDEPCTPKRGYKVHDEGHFAGSSVSFSCD 998  
 QY 999 PGVSLGSEELCLSEGERRTWDRPLTCVAECGCTVARGESQVLSPGIPAYEHNANCT 1058  
 DB 1247 PGVSLGSEELCLSEGERRTWDRPLTCVAECGCTVARGESQVLSPGIPAYEHNANCT 1058  
 QY 1059 WTEAEAGCTTIGHLVPDTEEVHDLIIMDGFVESGVLLKELSGPALPKDLHSTFNSV 1118  
 DB 1307 WTEAEAGCTTIGHLVPDTEEVHDLIIMDGFVESGVLLKELSGPALPKDLHSTFNSV 1118  
 QY 1119 LQSDPFTSKOGFALQFVSSTNATSGNDGICPONGSRSDSWEADGTVFQCDPGYALQ 1178  
 DB 1367 LQSDPFTSKOGFALQFVSSTNATSGNDGICPONGSRSDSWEADGTVFQCDPGYALQ 1178  
 QY 1179 SAEISCVKLENFFWQSPPTCIAPCGDLTGPSGVTILSPNPEEYPGKCEMDMVTYSP 1238  
 DB 1427 SAEISCVKLENFFWQSPPTCIAPCGDLTGPSGVTILSPNPEEYPGKCEMDMVTYSP 1238  
 QY 1239 DVIYALVNIENLEPGYDLFIATYDGRDLSPLIGSFYGSQULPGRJESSNSLFLAFRDA 1298  
 DB 1487 DVIYALVNIENLEPGYDLFIATYDGRDLSPLIGSFYGSQULPGRJESSNSLFLAFRDA 1298  
 QY 1487 DVIYALVNIENLEPGYDLFIATYDGRDLSPLIGSFYGSQULPGRJESSNSLFLAFRDA 1487

QY 1299 SVSNAQFVIDYTEMRESCFDPGSIKNGTRVGSDDLKLGSSVITYYCHGYEVEGTSTLSCI 1358  
 DB 1547 SVSNAQFVIDYTEMRESCFDPGSIKNGTRVGSDDLKLGSSVITYYCHGYEVEGTSTLSCI 1358  
 QY 1359 LGPDGKPVNNRPVCTATACGGQYVSDGVLSNPNPQYITSGOICILYVTVPKDXYVFG 1418  
 DB 1607 LGPDGKPVNNRPVCTATACGGQYVSDGVLSNPNPQYITSGOICILYVTVPKDXYVFG 1418  
 QY 1419 QPAFHTALNDVENVGDHSGHRSLLSLSGHSGESJPLATSNQVLIKFSKGLAPANG 1478  
 DB 1667 QPAFHTALNDVENVGDHSGHRSLLSLSGHSGESJPLATSNQVLIKFSKGLAPANG 1478  
 QY 1479 FHFVYQAVRSTATQCSSVPEERYKRLGSDFSVGAIVAREFECNSGVALQGSDEICLPVP 1538  
 DB 1727 FHFVYQAVRSTATQCSSVPEERYKRLGSDFSVGAIVAREFECNSGVALQGSDEICLPVP 1538  
 QY 1539 GALAONVAPTCVPCCGNLTERRGTTILSPFPPRYLNSLNCWKIYVBEAGIQIOVY 1598  
 DB 1787 GALAONVAPTCVPCCGNLTERRGTTILSPFPPRYLNSLNCWKIYVBEAGIQIOVY 1598  
 QY 1599 SEVTEQNDLSLEVPDADNTVYMLGSPGTTVPALINSTNOLYHAFYSDISVSAAGFHL 1658  
 DB 1847 SEVTEQNDLSLEVPDADNTVYMLGSPGTTVPALINSTNOLYHAFYSDISVSAAGFHL 1658  
 QY 1659 EYKTYGSLSCPEPAPVPSNGVKTGERLYLVNDVVSFOCEPVALQGHAIISCMPTVRRWNY 1718  
 DB 1907 EYKTYGSLSCPEPAPVPSNGVKTGERLYLVNDVVSFOCEPVALQGHAIISCMPTVRRWNY 1718  
 QY 1719 PPLCTIACCGTVEEMEGVILSPGPGVYPSNMDCSMTALPVGCAHIOCLNFTSEBNH 1778  
 DB 1967 PPLCTIACCGTVEEMEGVILSPGPGVYPSNMDCSMTALPVGCAHIOCLNFTSEBNH 1778  
 QY 1779 DYIEIRNGPYETSRMNGRFGSSSELPSLSTSHETTYVFNHSDSONRPGFLYQAVELQ 1838  
 DB 2027 DYIEIRNGPYETSRMNGRFGSSSELPSLSTSHETTYVFNHSDSONRPGFLYQAVELQ 1838  
 QY 1839 ECPDEPEPANGIVAGAGVNGQSTFECLPGYOLTHGVYLTCCQGTNRNMWHPKCEVP 1898  
 DB 2087 ECPDEPEPANGIVAGAGVNGQSTFECLPGYOLTHGVYLTCCQGTNRNMWHPKCEVP 1898  
 QY 1899 CGGNITSSNGTVSPGPFSSQDCWMLITVPDIGHVRLNLSLLOTEPBGDPFTIWDG 1958  
 DB 2147 CGGNITSSNGTVSPGPFSSQDCWMLITVPDIGHVRLNLSLLOTEPBGDPFTIWDG 1958  
 QY 1959 PQOAPRIGVTRSMAKTVQSSNOVYLKFRHDAATGIPALAFSAVPLTKCPPTILP 2018  
 DB 2207 PQOAPRIGVTRSMAKTVQSSNOVYLKFRHDAATGIPALAFSAVPLTKCPPTILP 2018  
 QY 2019 NAEVTEBENFENICDIYRVRCLPGFTLVGNEILLTCKLGYLOFEGRPPICEVHCFTELL 2078  
 DB 2267 NAEVTEBENFENICDIYRVRCLPGFTLVGNEILLTCKLGYLOFEGRPPICEVHCFTELL 2078  
 QY 2079 TDSGTGVLISQSPGSPYPOFTQCSWLVAREPDYNISLTVFELSKQDEFEIFDPGSPGQS 2138  
 DB 2327 TDSGTGVLISQSPGSPYPOFTQCSWLVAREPDYNISLTVFELSKQDEFEIFDPGSPGQS 2138  
 QY 2139 PLTKALSGVNSAPLIYVSSNSVYLRMSDHAANKRKFRIYAPYCSLPRAPLHGFTLG 2198  
 DB 2387 PLTKALSGVNSAPLIYVSSNSVYLRMSDHAANKRKFRIYAPYCSLPRAPLHGFTLG 2198  
 QY 2199 QTSSTOPGSGIHFQGNAGVRLVGHSMALICTRHPOGYHLSMAIPLCOALSGLBEAPNGM 2258  
 DB 2447 QTSSTOPGSGIHFQGNAGVRLVGHSMALICTRHPOGYHLSMAIPLCOALSGLBEAPNGM 2258  
 QY 2259 VFGKEVYVGTKAVYSCEGYHLQAGABATLACDLTGLMSNRNVPOCPVPTCDVSSISV 2318  
 DB 2507 VFGKEVYVGTKAVYSCEGYHLQAGABATLACDLTGLMSNRNVPOCPVPTCDVSSISV 2318  
 QY 2319 EHGKRWLIFETQYQOFOALMLICDPRGYTTCGVIVIFCOANGKSLSDSPTCTIISCGEL 2378  
 DB 2567 EHGKRWLIFETQYQOFOALMLICDPRGYTTCGVIVIFCOANGKSLSDSPTCTIISCGEL 2378  
 QY 2379 PIPNGHRIGTLISVYATATIFSCNSGTYLVGSRVRECMANGLMSGSEVCLACHGCTPBP 2438

```

Db 2627 SFPNGKIGTLVYGAATAFTCTGTGTVGSHVRECLANGMSSEFRCLAGHCGSDP 2686
Qy 2439 IYNGHIGENYSGVSYVYQCNAGFRILGMSVRICQDDHMSGKTPCVPTTCGPGNV 2498
Db 2687 IYNGHIGSDGSDYSDTYVYQCNPGFRVLGTSVRICLDHKKMSGOTPVCPVPTTCGPNPA 2746
Qy 2499 NGLGQSGQPNLVNPFVPCNGVYAEAGARSGCLASGQMSDMLTCRIINTDPGHENS 2558
Db 2747 HGFNNGSEFNLDVNFCTCNGTYLLQGVSRQCSNGMSPLTCRVMNSDGFVENA 2806
Qy 2559 VROVHAGSPHAFSGTIVSYRCHNGFYLLGTPVLSGCGDGTWDRBPQCLLVSCGHPSP 2618
Db 2807 IRHQQGNFSPSEFXGMSILYHCKKGFHLIGSSALTCWANGIMDRSLPKCLASGHPGV 2866
Qy 2619 PHQMSDSTTVAVVYRSCIGKRTLVGNSTRMCGLDGHWGSLPHSGTSVYCGPGI 2678
Db 2867 ANAVLTGELFTYGVAVVYHSCGSESLIGNDTRVCOBDSHWSGALPHCTGNNGPFCGPGT 2926
Qy 2679 PAHGIRLGDSPDPTVWRFSCEAGHVLRGSEERTCOANGMSGSGQPCGVISCGNPGPS 2738
Db 2927 PAHOSRLGDPTKSLRFSCEMCHOLRGSEERTCLANGMSGLQVCEAVSCNPGTPT 2986
Qy 2739 NARVVFSDGLVFSSSIYECREGYATGLSRHCSVNGTWTGSDPECLVINCDDPGI 2798
Db 2987 NGMIVSDGLIFSSSVIYACMEGYKTSGLMTRHCTANGTWTGTDCTIISCGDPTLAN 3046
Qy 2799 GLRLGNDPRNKTYTTCVPCGYMESHRSVSLCTKORTWNGIKPVCKALMKCPPLIPN 2858
Db 3047 GIDGTFTFNKITYSCONPBYVMAEATVATIRCTKGRMNPSPKPVKAVLCPPPPVON 3106
Qy 2859 GKVVGSDPFMWSSTVYACLEGYOLSLPAVPCGNGSGMTGELPOCFVFGSDPVPBSGR 2918
Db 3107 GYVEGSDPRMGSSISYSCMDYQLSHSHILSCEBRGWKGTIPQCLPFVCGDPEIPMBGR 3166
Qy 2919 REDRGFSYRSVSFSPCHPPLVGVSPRRFCQSDGTWSTGTPSCIDPTLTTCADPGVQFG 2978
Db 3167 LSGSPFYKSEVPFQCKSPFLVGVSSRRVOCADGTWGTGIDPTCIDPAHNTCPDGTHTFG 3226
Qy 2979 IONNSQGYGVSTVLFRCQKQYLLQGSTTRTCLPNLTWSTGTPDPCVPHHCRQPTPHAN 3038
Db 3227 IONSSRGYEVGVSTVFFRCRKGHYIQGSTTRTCLANLTWSTGIOTECIPHACROPTPHAD 3286
Qy 3039 VGALDLPMSGYTLI 3052
Db 3287 VRAIDLPTFGYTLV 3300

```

```

RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaïdo I., Osato N., Saito R., Suzuki H., Yamanka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schombach C., Gotojiori T.,
RA Baldairelli R., Hill D.P., Butt C., Hume D.A., Quackebush J.,
RA Schirral L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chochoia C., Coriani L.B., Cousine S.,
RA Dala E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimonard S., Gustincich S., Hirokawa N., Jackson I.U., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Meglert D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Partea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid Y., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Matanabe Y., Wells C.,
RA Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiroki T., Waki K., Kawai J., Aizawa K., Arikawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Ysuanishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Biney E., Hayashizaki Y.,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=0923L3-1; Sequence=Displayed;
CC Name=2;
CC IsoId=0923L3-2; Sequence=VSP 009037;
CC Note=No experimental confirmation available;
CC -1- SIMILARITY: Contains 14 CUB domains;
CC -1- SIMILARITY: Contains 28 Sush1 (CCP/SCR) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; AY017475; AAC54083.1; -
DR EMBL; AK038679; BAC30095.1; ALT_INTR.
DR EMBL; AK082377; BAC38482.1; -
DR HSP; Q9UCV4; INZI.
DR MGD; MGI:213783; Camd1.
DR InterPro; IPR000859; CUB1.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00431; CUB; 14.
DR Pfam; PF00084; Sush1; 28.
DR PROSITE; PS01180; CUB; 14.
DR PROSITE; PS50923; SUSH1; 28.
KW Alternative splicing; Repeat; Signal; Sush1; Transmembrane.
FT SIGNAL 1 29
FT CHAIN 30 3564
FT DOMAIN 30 3487
FT TRANSMEM 3488 3508
FT DOMAIN 3509 3564
FT DOMAIN 32 140
FT DOMAIN 143 204
FT DOMAIN 208 312

```

| FT   | DOMAIN   | 347  | 408  | Sushi 2.               |
|--|----------|------|------|------------------------|
| FT <th>DOMAIN</th> <td>411</td> <td>522</td> <td>CUB 3.</td>                     | DOMAIN   | 411  | 522  | CUB 3.                 |
| FT <th>DOMAIN</th> <td>525</td> <td>582</td> <td>Sushi 3.</td>                   | DOMAIN   | 525  | 582  | Sushi 3.               |
| FT <th>DOMAIN</th> <td>584</td> <td>692</td> <td>CUB 4.</td>                     | DOMAIN   | 584  | 692  | CUB 4.                 |
| FT <th>DOMAIN</th> <td>695</td> <td>756</td> <td>Sushi 4.</td>                   | DOMAIN   | 695  | 756  | Sushi 4.               |
| FT <th>DOMAIN</th> <td>758</td> <td>866</td> <td>CUB 5.</td>                     | DOMAIN   | 758  | 866  | CUB 5.                 |
| FT <th>DOMAIN</th> <td>871</td> <td>928</td> <td>Sushi 5.</td>                   | DOMAIN   | 871  | 928  | Sushi 5.               |
| FT <th>DOMAIN</th> <td>930</td> <td>1040</td> <td>CUB 6.</td>                    | DOMAIN   | 930  | 1040 | CUB 6.                 |
| FT <th>DOMAIN</th> <td>1043</td> <td>1102</td> <td>Sushi 6.</td>                 | DOMAIN   | 1043 | 1102 | Sushi 6.               |
| FT <th>DOMAIN</th> <td>1104</td> <td>1212</td> <td>CUB 7.</td>                   | DOMAIN   | 1104 | 1212 | CUB 7.                 |
| FT <th>DOMAIN</th> <td>1215</td> <td>1275</td> <td>Sushi 7.</td>                 | DOMAIN   | 1215 | 1275 | Sushi 7.               |
| FT <th>DOMAIN</th> <td>1277</td> <td>1386</td> <td>CUB 8.</td>                   | DOMAIN   | 1277 | 1386 | CUB 8.                 |
| FT <th>DOMAIN</th> <td>1389</td> <td>1449</td> <td>Sushi 8.</td>                 | DOMAIN   | 1389 | 1449 | Sushi 8.               |
| FT <th>DOMAIN</th> <td>1451</td> <td>1559</td> <td>CUB 9.</td>                   | DOMAIN   | 1451 | 1559 | CUB 9.                 |
| FT <th>DOMAIN</th> <td>1562</td> <td>1623</td> <td>Sushi 9.</td>                 | DOMAIN   | 1562 | 1623 | Sushi 9.               |
| FT <th>DOMAIN</th> <td>1625</td> <td>1733</td> <td>CUB 10.</td>                  | DOMAIN   | 1625 | 1733 | CUB 10.                |
| FT <th>DOMAIN</th> <td>1739</td> <td>1800</td> <td>Sushi 10.</td>                | DOMAIN   | 1739 | 1800 | Sushi 10.              |
| FT <th>DOMAIN</th> <td>1802</td> <td>1910</td> <td>CUB 11.</td>                  | DOMAIN   | 1802 | 1910 | CUB 11.                |
| FT <th>DOMAIN</th> <td>1913</td> <td>1972</td> <td>Sushi 11.</td>                | DOMAIN   | 1913 | 1972 | Sushi 11.              |
| FT <th>DOMAIN</th> <td>1974</td> <td>2082</td> <td>CUB 12.</td>                  | DOMAIN   | 1974 | 2082 | CUB 12.                |
| FT <th>DOMAIN</th> <td>2085</td> <td>2144</td> <td>Sushi 12.</td>                | DOMAIN   | 2085 | 2144 | Sushi 12.              |
| FT <th>DOMAIN</th> <td>2146</td> <td>2257</td> <td>CUB 13.</td>                  | DOMAIN   | 2146 | 2257 | CUB 13.                |
| FT <th>DOMAIN</th> <td>2256</td> <td>2317</td> <td>Sushi 13.</td>                | DOMAIN   | 2256 | 2317 | Sushi 13.              |
| FT <th>DOMAIN</th> <td>2319</td> <td>2430</td> <td>CUB 14.</td>                  | DOMAIN   | 2319 | 2430 | CUB 14.                |
| FT <th>DOMAIN</th> <td>2430</td> <td>2492</td> <td>Sushi 14.</td>                | DOMAIN   | 2430 | 2492 | Sushi 14.              |
| FT <th>DOMAIN</th> <td>2493</td> <td>2554</td> <td>Sushi 15.</td>                | DOMAIN   | 2493 | 2554 | Sushi 15.              |
| FT <th>DOMAIN</th> <td>2555</td> <td>2619</td> <td>Sushi 16.</td>                | DOMAIN   | 2555 | 2619 | Sushi 16.              |
| FT <th>DOMAIN</th> <td>2620</td> <td>2677</td> <td>Sushi 17.</td>                | DOMAIN   | 2620 | 2677 | Sushi 17.              |
| FT <th>DOMAIN</th> <td>2678</td> <td>2735</td> <td>Sushi 18.</td>                | DOMAIN   | 2678 | 2735 | Sushi 18.              |
| FT <th>DOMAIN</th> <td>2736</td> <td>2793</td> <td>Sushi 19.</td>                | DOMAIN   | 2736 | 2793 | Sushi 19.              |
| FT <th>DOMAIN</th> <td>2794</td> <td>2856</td> <td>Sushi 20.</td>                | DOMAIN   | 2794 | 2856 | Sushi 20.              |
| FT <th>DOMAIN</th> <td>2857</td> <td>2914</td> <td>Sushi 21.</td>                | DOMAIN   | 2857 | 2914 | Sushi 21.              |
| FT <th>DOMAIN</th> <td>2918</td> <td>2975</td> <td>Sushi 22.</td>                | DOMAIN   | 2918 | 2975 | Sushi 22.              |
| FT <th>DOMAIN</th> <td>2976</td> <td>3034</td> <td>Sushi 23.</td>                | DOMAIN   | 2976 | 3034 | Sushi 23.              |
| FT <th>DOMAIN</th> <td>3035</td> <td>3094</td> <td>Sushi 24.</td>                | DOMAIN   | 3035 | 3094 | Sushi 24.              |
| FT <th>DOMAIN</th> <td>3095</td> <td>3152</td> <td>Sushi 25.</td>                | DOMAIN   | 3095 | 3152 | Sushi 25.              |
| FT <th>DOMAIN</th> <td>3153</td> <td>3210</td> <td>Sushi 26.</td>                | DOMAIN   | 3153 | 3210 | Sushi 26.              |
| FT <th>DOMAIN</th> <td>3214</td> <td>3272</td> <td>Sushi 27.</td>                | DOMAIN   | 3214 | 3272 | Sushi 27.              |
| FT <th>DOMAIN</th> <td>3273</td> <td>3332</td> <td>Sushi 28.</td>                | DOMAIN   | 3273 | 3332 | Sushi 28.              |
| FT <th>DOMAIN</th> <td>40</td> <td>40</td> <td>N-linked (GlcNAc. . .)</td>       | DOMAIN   | 40   | 40   | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>57</td> <td>57</td> <td>(Potential).</td>               | CARBOHYD | 57   | 57   | (Potential).           |
| FT <th>CARBOHYD</th> <td>587</td> <td>587</td> <td>N-linked (GlcNAc. . .)</td>   | CARBOHYD | 587  | 587  | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>686</td> <td>686</td> <td>(Potential).</td>             | CARBOHYD | 686  | 686  | (Potential).           |
| FT <th>CARBOHYD</th> <td>955</td> <td>955</td> <td>N-linked (GlcNAc. . .)</td>   | CARBOHYD | 955  | 955  | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>1015</td> <td>1015</td> <td>(Potential).</td>           | CARBOHYD | 1015 | 1015 | (Potential).           |
| FT <th>CARBOHYD</th> <td>1034</td> <td>1034</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 1034 | 1034 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>1184</td> <td>1184</td> <td>(Potential).</td>           | CARBOHYD | 1184 | 1184 | (Potential).           |
| FT <th>CARBOHYD</th> <td>1197</td> <td>1197</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 1197 | 1197 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>1399</td> <td>1399</td> <td>(Potential).</td>           | CARBOHYD | 1399 | 1399 | (Potential).           |
| FT <th>CARBOHYD</th> <td>1454</td> <td>1454</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 1454 | 1454 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>1572</td> <td>1572</td> <td>(Potential).</td>           | CARBOHYD | 1572 | 1572 | (Potential).           |
| FT <th>CARBOHYD</th> <td>1644</td> <td>1644</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 1644 | 1644 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>1792</td> <td>1792</td> <td>(Potential).</td>           | CARBOHYD | 1792 | 1792 | (Potential).           |
| FT <th>CARBOHYD</th> <td>1805</td> <td>1805</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 1805 | 1805 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>1882</td> <td>1882</td> <td>(Potential).</td>           | CARBOHYD | 1882 | 1882 | (Potential).           |
| FT <th>CARBOHYD</th> <td>2018</td> <td>2018</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 2018 | 2018 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>2149</td> <td>2149</td> <td>(Potential).</td>           | CARBOHYD | 2149 | 2149 | (Potential).           |
| FT <th>CARBOHYD</th> <td>2154</td> <td>2154</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 2154 | 2154 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>2187</td> <td>2187</td> <td>(Potential).</td>           | CARBOHYD | 2187 | 2187 | (Potential).           |
| FT <th>CARBOHYD</th> <td>2358</td> <td>2358</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 2358 | 2358 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>2394</td> <td>2394</td> <td>(Potential).</td>           | CARBOHYD | 2394 | 2394 | (Potential).           |
| FT <th>CARBOHYD</th> <td>2400</td> <td>2400</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 2400 | 2400 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>2445</td> <td>2445</td> <td>(Potential).</td>           | CARBOHYD | 2445 | 2445 | (Potential).           |
| FT <th>CARBOHYD</th> <td>2470</td> <td>2470</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 2470 | 2470 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>2503</td> <td>2503</td> <td>(Potential).</td>           | CARBOHYD | 2503 | 2503 | (Potential).           |
| FT <th>CARBOHYD</th> <td>2605</td> <td>2605</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 2605 | 2605 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>2750</td> <td>2750</td> <td>(Potential).</td>           | CARBOHYD | 2750 | 2750 | (Potential).           |
| FT <th>CARBOHYD</th> <td>2761</td> <td>2761</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 2761 | 2761 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>2795</td> <td>2795</td> <td>(Potential).</td>           | CARBOHYD | 2795 | 2795 | (Potential).           |
| FT <th>CARBOHYD</th> <td>2894</td> <td>2894</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 2894 | 2894 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>2963</td> <td>2963</td> <td>(Potential).</td>           | CARBOHYD | 2963 | 2963 | (Potential).           |
| FT <th>CARBOHYD</th> <td>3022</td> <td>3022</td> <td>N-linked (GlcNAc. . .)</td> | CARBOHYD | 3022 | 3022 | N-linked (GlcNAc. . .) |
| FT <th>CARBOHYD</th> <td>3056</td> <td>3056</td> <td>(Potential).</td>           | CARBOHYD | 3056 | 3056 | (Potential).           |

```
Db 1067 LAFPCFGYRLLEGATKLTCLCGGRVMSAPLPRCAECGASVKNBEGTLLSPNPSHYDN 1126
Qy 880 NHECIYSIQTPGKGIQAKAPAFELSEEDVLKATDGNNSAKLLGVSHSMEGVTLLNST 939
Db 1127 NHECIYIETBAGKIHRLARTFQLFEGDTLKVYDGKSSRSRGVFRSSEFMGLVJNST 1186
Qy 940 SSSLMLPITDAENTSKGFLHSSFELIKCEDPCTPKFGYKHYDEGHFACSSVSFQCDP 999
Db 1187 SNYLRLEPNYNGSTTAQGFOLTYTSFDLVKCEDPEGIPMYGRIKDDGHFTDVTVLYSCNP 1246
Qy 1000 GYSLRGSEBELLCSEGERRTMDRPLPTCAECGTVRGESGVLSPGYPAYEHLNLCIW 1059
Db 1247 GYAHGSGSTLTLCSGDRRWMDKPMPSVCAECGLVHAATSGRIILSPGPAYDNNHCTW 1306
Qy 1060 TIEBAGCTIGLHLVLPDTEBHVLRKWGPVSGVLKELSGPALPKDLHSTFNSVYL 1119
Db 1307 TIEADPKITSLHIVDTEHMDILKWDGPDVSNLLKEMSGSALPEDIHSTFNSLTL 1366
Qy 1120 QFSTDPFTSKGFALIOFSVSTATSCNDPGIPQNGSRSDSWEAGDSTVFOCDPGVALQGS 1179
Db 1367 QFSDPFTSKGSFSTQFSTSLASTCNDPMQNGRTYGDSEBDDTTFOCDPGIOLQGP 1426
Qy 1180 AELSCVKIENRFWQPSPTICIAFCGGDLTGPSGVILSPNYBEYPGKECDMAKVTYSPD 1239
Db 1427 AKICVQLNRRFFQOPBPSCIIAOGGDLTGPAVILSPNYPOYRPGKECDMAKIXNPD 1486
Qy 1240 YVIALVNIENFLECYDFLHLYDGRDSLPLISGFYSGQLPGRIESSNSLFLAFSDAS 1299
Db 1487 FVIALIFKFSFMEBSPYFLHLYEGEDSNSPLISGFQCSQAPERIESSGNSLFLAFSDAS 1546
Qy 1300 VSNAGFIIDYENRESQFDPGSIKNGTRVGSDLKLSVTVYHGHGVGEGTSTLSL 1359
Db 1547 VGSGLPIERKEKREACFPDGNIMNGRICTDFLGSVTYQCDSGKTIWPSISCVT 1606
Qy 1360 GPDPKPYMNNRPVCTAPCCGQYVGSQDVVLSPNYPQNYTSGQICLYFVVPKDYVVFQ 1419
Db 1607 GADGKPSMDRALPACQAPCCGQYVGSQDVVLSPNYPHNYTAGQMCYVSIYTKXKFFVQ 1666
Qy 1420 PAFHTILANDVVEVHDSHSHSLSSLSGSHTEBSLPLATSNQVLIKFSAGLAPRGF 1479
Db 1667 FAYFOTILNDLAEFLDGTHPQARLLSSLSGSHSGETPLATSNQVLIKFSAGSARGF 1726
Qy 1480 HFVYQAPRTSATQCSSVPEBRYGRLGSDPSVGAIVRECNAGVALQGSSEIECLVPBG 1559
Db 1727 HFVYQAPRTSDTCSSVPEBRYGRIGSEFSAGSIVRECPGTYLLQGSTAIRQOSVBN 1786
Qy 1540 ALAQMNVSAPTQVPCGQNLTERGTLLSPGPPEPYLNSLNCVWKIYVBEAGIQIOVNS 1599
Db 1787 ALAQMNVTIPSCVPCGQNLTERGTLLSPGPPEPYLNSLNCVWKIYVBEAGSQIOVNS 1846
Qy 1600 FVTEQWMSLEVPDADNTVYMLGSPSGTVPALLNSTSNQVLIHFSYDISVSAAHFLB 1659
Db 1847 FATEQWMSLEIHDGDMTAPRLGSPGTVPALNSTSNQVLIHFSYDISVSAAHFLB 1906
Qy 1660 YKTVGLSSCEBPAPNSGVKTGERLYLVNDVVSFOCEEGYALQGAHISCMPTYRRMNYR 1719
Db 1907 YKTVGLAACBPAPNSGKIGDRYVNDVLSFOCEEGYALQGSISHSCMPTYRRMNYR 1966
Qy 1720 PPLCIAOCCGTVBEMEGVILSPGPGNYPSNMDCSMKIALPVGGAHIOFLANFTEBPNH 1779
Db 1967 SPCLIAOCCGTVBEMEGVILSPGPGSYPNNDCTWKIISLPVIGGAHIOFLANFTEBPNH 2026
Qy 1780 YIEIRNCPYETSRMGRFSGSEBPSLSLSTSHETTVYFHSDSHONRPGFKLEYAYELOE 1839
Db 2027 YLIEIQNRPYHSSPMMQFSGPDLPTSLSTHETTLIRFYSDHSQNRQGFKLXYAYELOE 2086
Qy 1840 CPDPEEPANGIVRAGVNGQSVTFECLPQVLGHLVLTQGHSTNNMHPKCEVPC 1899
Db 2087 CPDPEEPANGFMINDSDSVGQISIFECYPGYILGHLVLTQGHSTNNMHPKCEVPC 2146
Qy 1900 GGNITSNGTVYSGFSPSSQDCVWLITVPYIGHGVLNLSLQTEPSGDFITINDGP 1959
Db 2147 GYNTVSQNGTVYSGFPEDEXPILEKDCMLVTVPPGHGVYINFLLQTEAVNDYIAVMDGP 2206
Qy 1960 QQTAPRLGVTRSAKTKVQSSSNQVLLKTHRDAAATGIPALASAPLTKCPPTLLPN 2019
Db 2207 DQNSPQGVSPGNAPEBTASSTNQVLLKTHSDPSNNGGFVLANHAOLKRCPPPPVPO 2266
Qy 2020 AEVVTENEEFNIGIVAYRCILPGFTLVGNELTKLGTVLOFEGPPRICEVHCTNELLT 2079
Db 2267 ADLITEBDEFEIDFVYKQCHPGFTLLGSDTLTKKLSQLLFQSPSPPTCEACQANEVRT 2326
Qy 2080 DSTGVILSOSYPSGFOTCSMLVRYVEPDYNI SLTYEYFLSEKODEFEI FDGSPGQSP 2139
Db 2327 ESSGVILSPGPGNYFNSQCAWGIKYPMPNITLFDVTQSKQFPALEVFQSSGRSP 2386
Qy 2140 LKALSGNYGAPLIVSSSSVYLRMSSDHAYNRKGRKIRYSAPYCGLPAPLHGFLLQ 2199
Db 2387 LVLVLSNHTHEQSNFTSRSHNLYLRWSTDHATSKKGFKIRYAAVYCSLTSLRNGGLNK 2446
Qy 2200 TSOTPGSIFPGCAGYRLVGHSMALCTRRPOGYHLSEBAPILCQALSCGLPEAPKXGMV 2259
Db 2447 TAGAVSKAHYFCKPGRMTGHSNATCRNRPVGVYQMDSNAFLCQAVSCGIPBAPNGSF 2506
Qy 2260 FGKEYTVGTRAVYSCSEGYHLOAGAEATACDLTGLSNBNRPQCVVTCPPVSSISVE 2319
Db 2507 TGNEFTLDSKVTYECNEGFLDASQEAATVQCEBDGLMSNKGKPTCKPVPQPSIEQULSE 2566
Qy 2320 HGRNRLIFETQYQFOALMLICDPGYTYTGQRYVIRCOANGKWSLGSSTPTCRILISCEBLP 2379
Db 2567 HVMRLVSGSLINEGAVLLSCSPGYFLQCRLLQCCANGTWSFEEDRPCKVYISCSLS 2626
Qy 2380 IIPNGHRIGLTVYGAATAPSCNSGYTLVGSRYRECMANGLMSGEVRCLAGHGTEPEPI 2439
Db 2627 FPPNGKIGTLITGATPALTCTNIGTLVSSHRECLANGLMSSETERCLAHGGSDDPI 2686
Qy 2440 VNGHINEANTSYRGSVYQCNAGFRLLGMSVRLICQODHMSGKTPFCVPTTCGHGPNVN 2499
Db 2687 VNGHISGDGSHYDVTYQCNPGFRVLGTVSVRLICRTTSRGRVLCVPTTCGHGPNVAH 2746
Qy 2500 GLTQGNQFNLDVYKFCVNGPYABEGARQCLASGQMSDMLPTCRILINCTDPGHQENSV 2559
Db 2747 GLTNGTEFNINDLVNFCVHGYRLOGASRAQCRSNGQMSPLPCCRVAVNSDPSGVNNAV 2806
Qy 2560 ROYVASGPHRFSGFTTYRCHNGFYLLGTPVLSQCGDGMWRDRPOCLVLSGHPSP 2619
Db 2807 RHQGNPESFEGTSMYHCKTGTFYLLGSSALTCMASGLMDSLPCLLAISGHPVPA 2866
Qy 2620 HSGWGSYTVGAVVRYSCTGKRTLVNGSTRMGLDGHMTGSLPHCSGTSVGVCGDGP 2679
Db 2867 NAVLTGSLFTYGATVQYSCAGGQILTGNSRTRVQEDBSHMGSLPHCGSNRPGFCGDGTP 2926
Qy 2680 AHGIRLGSDFDPTWRFSCBAGHVLRGSEERTCOANGSNVSGQPECGVITSCGNPGTSPN 2739
Db 2927 AHSGRILDEFEKTKSLRFSCEMGHQLRGFAERTCLVNGVSGVQPVCEAVSCGNPGTPTN 2986
Qy 2740 ARVFPSSGIVFSSSIVYECBEGYATGLSRHSVNGTMTGSPBECLVINCQPGIRANG 2799
Db 2987 GMLSSDGLIFSSSVIYACWEGYKTSGLMRHCTCANSTGTATDCTITISGPGTLPNG 3046
Qy 2800 LRLANDPFRYKTVTYOCVPGYMMESHRSVYLSCTKDRTNMGTPVCKALCKCPPLIPNG 2859
Db 3047 IQGTGDTFPMKTVSYQCNPGYLMERPPTSPIIRCTKGTQWQNSRPLCAVALCQNPBPVNG 3106
Qy 2860 KVVGSDFMGSSVTVACLBGTYLSLPAVFTCEBNGSVTGLPOCFPVYFGDPGVPSGRR 2919
Db 3107 KVGSDPRMWSASISYSCVDYQULSHSALISCEBGRVWKGAVPOCLPVFCBDPGTAPBGR 3166
Qy 2920 EDGFSYRSVSFSCHPRLVTVGSPPRFQSDGMSGTQSCIDPITLTADGVPOFGI 2979
Db 3167 SGKSFYKSEVFTQCKEPFVLVGSRRCTQADGIMSGIQTCTIDPAHTACPDGTPHFI 3226
Qy 2980 QNNSQGYOVSFVLFRCQKGYLLQGSTTRTCLPNTWSGTPPCVPHHCQPEPTTANV 3039
Db 3227 QNNSKGYEVGSYVFPFRCKKYNHIOGSTTRTCLANTLWMSGQTEBICPHACQPEPTPAADV 3286
```

Qy 3040 GALDIPSWGYTLI 3052  
 Db 3287 RAIDLPACGYTLV 3299  
 RESULT 4  
 CSM3\_HUMAN  
 ID\_CSM3\_HUMAN STANDARD, PRT, 3670 AA.  
 AC Q72407.096P23;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE CUB and sushi multiple domains protein 3 precursor.  
 GN Name=CSMD3; Synonyms=KIAA1894;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RX MEDLINE=22788796; PubMed=12906867; DOI=10.1016/S0888-7543(03)00149-6;  
 RA Ian W.L., Scholnick S.B.,  
 RT "Identification of two new members of the CSMD gene family.";  
 RL Genomics 82:412-415(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), TISSUE SPECIFICITY, AND  
 RC VARIANTS MET-182 AND HIS-3584.  
 RX TISSUE=Brain, and Testis;  
 RA PubMed=12943675; DOI=10.1016/S0006-291X(03)01555-9;  
 RX Shimizu A., Asakawa S., Shimizu N.,  
 RT "A novel giant gene CSMD3 encoding a protein with CUB and sushi  
 RT multiple domains: a candidate gene for benign adult familial myoclonic  
 RT epilepsy on human chromosome 8q23.3-q24.1.";  
 RL Biochem. Biophys. Res. Commun. 309:143-154(2003).  
 RN [3]  
 RP SEQUENCE OF 624-3670 FROM N.A. (ISOFORM 4), AND TISSUE SPECIFICITY.  
 RC TISSUE=Brain;  
 RX MEDLINE=21456161; PubMed=11572484;  
 RA Nagase T., Kikuno R., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human genes. XXI.  
 RT The complete sequences of 60 new cDNA clones from brain which code for  
 RT large proteins.";  
 RL DNA Res. 8:179-187(2001).  
 RN [4]  
 RP SEQUENCE OF 2234-3670 FROM N.A. (ISOFORM 5).  
 RC TISSUE=Liver;  
 RX PubMed=14702039; DOI=10.1038/rj1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
 RA Nagahori K., Murakami K., Yasuda T., Iwawaki T., Magatsuma M.,  
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
 RA Yamazaki M., Nishimura K., Iehibashi T., Yamashita H., Murakawa K.,  
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y.,  
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hattori T.,  
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
 RA Nomura Y., Togiya S., Komai F., Hara R., Takuchi K., Azita M.,  
 RA Imose N., Mueshino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
 RA Moriya S., Momiyama H., Satoh N., Takami S., Teraishi Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Teshiro H., Taniguchi A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,

RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.,  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 CC - SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC - ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=5;  
 CC Name=1;  
 CC IsoId=Q72407-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q72407-2; Sequence=VSP\_009047;  
 CC Name=3;  
 CC IsoId=Q72407-3; Sequence=VSP\_009048, VSP\_009049;  
 CC Name=4;  
 CC IsoId=Q72407-4; Sequence=VSP\_009050;  
 CC Note=No experimental confirmation available;  
 CC Name=5;  
 CC IsoId=Q72407-5; Sequence=VSP\_009051, VSP\_009052;  
 CC Note=No experimental confirmation available;  
 CC - TISSUE SPECIFICITY: Weakly expressed in most tissues, except in  
 CC brain. Expressed at intermediate level in brain, including  
 CC cerebellum, substantia nigra, thalamus, spinal cord, hippocampus  
 CC and fetal brain. Also expressed in testis.  
 CC - SIMILARITY: Belongs to the CSMD family.  
 CC - SIMILARITY: Contains 14 CUB domains.  
 CC - SIMILARITY: Contains 28 Sushi (CCP/SCR) domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AY210419; AAC034702.1; -;  
 DR EMBL; AB114604; BAC82444.1; ALT\_INIT.  
 DR EMBL; AB114605; BAC82444.1; -;  
 DR EMBL; AB067481; BAB67787.2; -;  
 DR EMBL; AK126252; BAC6505.1; ALT\_INIT.  
 DR Genew; HGNC:19291; CSMD3.  
 DR MIM: 608399; -;  
 DR Interpro; IPR000859; CUB.  
 DR Interpro; IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF00431; CUB; 14.  
 DR Pfam; PF00084; Sushi; 27.  
 DR PROSITE; PS01180; CUB; 14.  
 DR PROSITE; PS0923; SUSHI; 28.  
 KW Alternative splicing; Polymorphism; Repeat; Signal; Sushi;  
 KW Transmembrane.  
 FT SIGNAL 1 23  
 FT CHAIN 1 3670 Potential.  
 FT DOMAIN 24 3593 CUB and sushi multiple domains protein 3.  
 FT TRANSMEM 3594 3614 Extracellular (Potential).  
 FT DOMAIN 3615 3670 Potential.  
 FT DOMAIN 28 136 Cytoplasmic (Potential).  
 FT DOMAIN 139 200 Sushi 1.  
 FT DOMAIN 204 308 Sushi 2.  
 FT DOMAIN 447 508 Sushi 2.  
 FT DOMAIN 511 622 Sushi 3.  
 FT DOMAIN 625 682 Sushi 3.  
 FT DOMAIN 684 792 Sushi 4.  
 FT DOMAIN 795 856 Sushi 4.  
 FT DOMAIN 858 966 Sushi 5.  
 FT DOMAIN 971 1028 Sushi 5.  
 FT DOMAIN 1030 1140 Sushi 5.  
 FT DOMAIN 1143 1202 Sushi 6.  
 FT DOMAIN 1204 1312 Sushi 7.  
 FT DOMAIN 1315 1375 Sushi 7.  
 FT DOMAIN 1377 1486 Sushi 8.  
 FT DOMAIN 1489 1549 Sushi 8.

```

FT DOMAIN 1551 1659 CUB 9.
FT DOMAIN 1662 1723 Sush1 9.
FT DOMAIN 1725 1833 CUB 10.
FT DOMAIN 1839 1900 Sush1 10.
FT DOMAIN 1902 2010 CUB 11.
FT DOMAIN 2013 2072 Sush1 11.
FT DOMAIN 2074 2182 CUB 12.
FT DOMAIN 2185 2244 Sush1 12.
FT DOMAIN 2246 2357 CUB 13.
FT DOMAIN 2356 2417 Sush1 13.
FT DOMAIN 2419 2530 CUB 14.
FT DOMAIN 2530 2592 Sush1 14.
FT DOMAIN 2593 2654 Sush1 15.
FT DOMAIN 2655 2719 Sush1 16.
FT DOMAIN 2720 2777 Sush1 17.
FT DOMAIN 2778 2835 Sush1 18.
FT DOMAIN 2836 2893 Sush1 19.
FT DOMAIN 2894 2955 Sush1 20.
FT DOMAIN 2956 3013 Sush1 21.
FT DOMAIN 3017 3074 Sush1 22.
FT DOMAIN 3075 3133 Sush1 23.
FT DOMAIN 3134 3193 Sush1 24.
FT DOMAIN 3194 3251 Sush1 25.
FT DOMAIN 3252 3309 Sush1 26.
FT DOMAIN 3313 3371 Sush1 27.
FT DOMAIN 3372 3431 Sush1 28.
FT CARBOHYD 36 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 53 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 324 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 372 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 687 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 786 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 929 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1055 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1089 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1134 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1243 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1499 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1554 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1572 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1744 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1892 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 1982 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2118 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2249 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2254 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2287 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2458 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2500 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2574 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2705 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2825 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2895 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 2915 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 3062 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 3121 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 3130 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 3157 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 3171 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 3181 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 3239 N-linked (G1cNAc . . .) (Potential).
FT CARBOHYD 3327 N-linked (G1cNAc . . .) (Potential).

Query Match 62.4%; Score 10591.5; DB 1; Length 3670;
Best Local Similarity 60.1%; Pred. No. 0;
Matches 1816; Conservative 501; Mismatches 665; Indels 41; Gaps 6;

```

```

Db 471 VQFSCDEBYVLQGAASITCQRIAEVFAWSDHRRPVCKRYKTCGSLNLCQSGTFTSPNPFQ 530
156 YDNNACWWTITANPKVITKLAEEFDELRGYTTLVVGCGGDDQDKTYLVMQNCSD 215
531 YDSNAQCWWTAVANKVITQINEEFDELDGYTTLVVGCGEVDPRVYQV----- 583
216 SPHTPGRIPE---SMGDIWRQKVTALICRDISSDASGVRKPKTSNAVELVAPG 272
584 ---LTGSFVFDLVMSQMWLHQT-----DESQSVGFKNVK----- 620
273 TEIBQSCGDPGIPAYARRREGSRPHDGTKEFCQAPFELVQKAITCOKNNQMSAKPG 332
621 -EIKESGDPBGTLYGIRREGDGSNRDVALREFCQGFELIGEXIVCOENQMSANPI 679
333 CVESCFNFTSPSGVLTSPNRYEYGHILCWTILARPESRITLANNDIVDQRPV 392
680 CIFPCLSNFTAPMGTVLSPDPYEGYGNLNCIWTIISDPGRHLSPNDPDLSEQPF 739
393 IKDQATAEAPVLTGFSGNQLPSRITSSGHVAREFQTDHSTGKRGFNITFTPHNDCPD 452
740 VKQDSFESPLGTFTGAEPVSHLTSNHLRLEFQADHMSGNGFNITVTPHNECPD 799
453 PGVAVNGKRPDLSIQGSSISFLCDEGFLGTQSEITTVYLKEGSSVYMSAVLCEAPCG 512
800 PGIRINARRGDNFQLGSSISVICEBGFIKQGTETITCILMDGKVMWSGLIPKCAPCG 859
513 GHLTSPGTLSPMPGFYKDALSCAVIAEOPVPIKITPDRKTEVNTDLEVRGRT 572
860 GHFAPSGVILSPMPGYKDSLNCENVIAEPEHSHIKITERQOTELANDVLEHVGPN 919
573 YSAPLIVYNGTQVPOPLISTSNLYLFFSTDKSHSDIGFOLRJETITLSDHCLDGIP 632
920 LLSPLLSYNGTQVPOPLFSSSNFYLLFTTDSRSNNGKRIHESVTVNTVSLDGP 979
633 VNGQRHNDYVYALVTFSCDSGYTSLDGBELCEPFPQWSRALPSCALCGFIQSSG 692
980 VHGRRYGHDSISGTVFSCDSGYRLSHEPPLCEKHMMSHPLPTCDALCGGVDRPSG 1039
693 TILSPGPDYVYNNLCTWILLETSHGKGVFTFTFTFLBSGHVLLTTENGSPFOPLRQ 752
1040 TILSPGPEYVYNNLCTWIVDVYHGKGVFNFTFLBHDHDLTTENGSPFOPLRQ 1099
753 TGSRLPAPISAGYNGTAAQVRFISDPSMEGNIIFSEXDLEPCEPEVPAVIRKL 812
1100 TGSRLPPTIAGLYGNRAQRFISDPSISYEGFNITFSEYNLEPCDDPGIPQSGSITG 1159
813 QPVGVDLTTFSCPEGRVLEGTARITCLGGRRLMSPLPRCVAECSNVTGQTLSPN 872
1160 NFGIGDILTFSCSGYRLBEGSEITCLGGRRWASAPLPRCVAECSATNNEGILLSPN 1219
873 FPNVYNNHNCIYSIQOPKGIOLKARAELEBGDVLKYDDGNNSARLLGVSHSEMM 932
1220 YPLAVYNNHNCIYSIQOAKGKINISARTFHLAQGDVLKYDDKDTHTLLGAVTASMR 1279
933 GVTLNYSSTSLMTDFITDANTSQFELHSFELICEDPGMPKFGYKXVDESHFAGSS 992
1280 GLTISSTSNQWMEFNSDTEGIDGFLVTSFELSCBEPGLRPFQYKISDQHFAGST 1339
993 VSPSCDPYSLRGSEBELCLSGERRTDRPLPTCVABCGSTVNGEVSQVLSPPYAPYE 1052
1340 IYGCNNGYTLHSGSLKCMGTGERRAMDYPLPSCIAACGGRFKGESGRILSPYPPYD 1399
1053 HNLNCTITAEAGCTIGLHFLYFDTEBVHDVLRINQGVYSGVLLKELSGPALPKLHS 1112
1400 NNLRCMMIMIVDPENIYSLQFLAFDTASHDILRVMDGPPENMDLLEISGSLIPRGIHS 1459
1113 TPNVSVLPQSDPFTSQGFALQPSVSTATSNCDPGLPQNGSGSDGSEWAGDSTVFPQCDP 1172
1460 TNLVITVITQFDITFYSKGFALQPSVSTATSNCDPGLPQNGSGSDGSEWAGDSTVFPQCDP 1519
1173 GVALQGSABEISCVKIEENRFPWQSPPTCIAPCGGDLTGPSGVLLSPNYPAPYPGKCDW 1232
1520 GYELQGERITTCIQENRFRWQSPPTCIAPCGGDLTGPSGVLLSPNYPAPYPHYSRCDW 1579

```

QY 1233 KTVSPDVIALVFNIFNLEPGYDPLHIYDGRDLSPLIGSFYSGOLPGRTESSNSLFL 1292  
 DB 1580 TITVADYVIALAFISFISERNYDLIYDGPDSNLSPLIGSFODSKLEPERLESSNTMHL 1639  
 QY 1233 AFRSASVSNAGFVIDYENPRESCEFDPSIKNGTRVSGDLKSSVTVYCHGGVEYECT 1352  
 DB 1640 AFRSGSVSYTGFFHLEKAKLRESCFDPGNIMNGTRIGMDKYLGSVTVYDAGAVIYQGY 1699  
 QY 1353 STLSCILGPDGKPVNNRPNVCTAPCGQGVYSGCVLSNPNYQNYTSGLCYFYVYPK 1412  
 DB 1700 STLTIMDDGPRGNRRLPSCNAPCSRSSTGSEGTVLSPNYPKNYSVGHNCVSIAPX 1759  
 QY 1413 DYVVGQFAPFHTLNDVVEYHDSHQSRLLSLSGSHTESEPLATSNQVLKFAKG 1472  
 DB 1760 EFVVGQFVFPQTSIHDVVEYDPTQSSLSLSGSHSESLPLSSGNQITIRFISVG 1819  
 QY 1473 LAPARGFHVYQAVPRTSATOCSSVPEPRYKRLGSDFSVGAIVFECNSGYALQSGPEI 1532  
 DB 1820 PITAKGFHVYQAVPRTSSTOCSSVPEPRYKRLGSDFSVGAIVFECNSGYALQSGPEI 1879  
 QY 1533 ECLPVPGLAQMNVAPTCVPCCGNLTERRGTLISGEPPEYLSNLCVWKIYVPEAG 1592  
 DB 1880 RCEVVPNSLAQMNLSPLCIYPCCGILTKKGTILSPGPEYDNNLNCVWKITVPEAG 1939  
 QY 1593 IQIVVSVFTVQNDLSLEVPDADNTVYMLGSFSGITVPALNLTSTNQLYHFEYSDISVS 1652  
 DB 1940 IQOVVSVFATEHNMDSLEFDYGDGNNAPRLGSYSGTTIIPHLNLTSTNNLVLNQSDISVS 1999  
 QY 1653 AAGHLEKTYVLSCEBPANVSNGVKTGERLYLNDVVSFQCEBGVYALQGHANISCPGT 1712  
 DB 2000 AAGHLEKTYVLSCEBPANVSNGVKTGERLYLNDVVSFQCEBGVYALQGHANISCPGT 2059  
 QY 1713 VRANVVPPLCIAQCGGTVEEMEGVILSPGPGNYPNMDCSMKIALPVFGAHIQFIN 1772  
 DB 2060 VRANVVPPLCIAQCGGTVEEMEGVILSPGPGNYPNMDCSMKIALPVFGAHIQFIN 2119  
 QY 1773 STEPNHYIETRNPRYTSRMGRFSGSELPSLSLSTSHETTVYFHSHSQNRPGKLEY 1832  
 DB 2120 STEPNHYIETRNPRYTSRMGRFSGSELPSLSLSTSHETTVYFHSHSQNRPGKLEY 2179  
 QY 1833 QAVELQCEPDEPFANGIVAGAGYNGVQSTFECLEPQYLTGHPVLTCHQTRNMDHL 1892  
 DB 2180 QAVELQCEPDEPFANGIVAGAGYNGVQSTFECLEPQYLTGHPVLTCHQTRNMDHL 2239  
 QY 1893 PKCEVPCCGNITSSNGTVSPGFPSPYSQDCVWLITVPIGHVRLNLSLQTEPSGDF 1952  
 DB 2240 PKCEVPCCGNITSSNGTVSPGFPSPYSQDCVWLITVPIGHVRLNLSLQTEPSGDF 2299  
 QY 1953 ITIWDGQOATAPRLGVFTRMAKKTQSSNOVLKFRHDATGCIPTAFSAVPLTKCP 2012  
 DB 2300 ITIWDGQOATAPRLGVFTRMAKKTQSSNOVLKFRHDATGCIPTAFSAVPLTKCP 2359  
 QY 2013 PPTILPNAEVVTEBENFNGIDIVRYRCLPGFTLVAGNEILTKGLYLOPESPPICEVHC 2072  
 DB 2360 PPTILPNAEVVTEBENFNGIDIVRYRCLPGFTLVAGNEILTKGLYLOPESPPICEVHC 2419  
 QY 2073 PTMELTLDSTGVLSQSYGSPYQFQTCMWLVREVDYNI SLTVEYFLESEKQYDEFEIFD 2132  
 DB 2420 PTMELTLDSTGVLSQSYGSPYQFQTCMWLVREVDYNI SLTVEYFLESEKQYDEFEIFD 2479  
 QY 2133 GPGSGPPLKALSGNSAPLITVSSNSVYLRSSPHANRKGFKIRYAPICSLPRAPL 2192  
 DB 2480 GPGSGPPLKALSGNSAPLITVSSNSVYLRSSPHANRKGFKIRYAPICSLPRAPL 2539  
 QY 2193 HGFILQISTOPGSIHFCGNAGYRLVGHSMALCTRHPOGYHLMSEAPLLOALSCGLPE 2252  
 DB 2540 HGFILQISTOPGSIHFCGNAGYRLVGHSMALCTRHPOGYHLMSEAPLLOALSCGLPE 2599  
 QY 2253 AFRNGVAFKEVTVGKAVYSCSEGYHLQAGAEATACLDGLWSNRNVPQCVPTCPD 2312  
 DB 2600 AFRNGVAFKEVTVGKAVYSCSEGYHLQAGAEATACLDGLWSNRNVPQCVPTCPD 2659

QY 2313 VSSISVEHGMRLIFETQIOFQAQMLICDPGYVYTGQARVIRCQANGKSLGSDTEPCRI 2372  
 DB 2660 INSFTILEHGMRLVNSHVEYKTKVFSQDPCGHGLGPASIECLPNGTSMWRNERPYCQI 2719  
 QY 2373 ISCGEPIPNNGRIETGLSVYGTALFSCNSGVTIVSGVREEMANGMLSGSVRLCLAH 2432  
 DB 2720 ISCGEPIPNNGRIETGLSVYGTALFSCNSGVTIVSGVREEMANGMLSGSVRLCLAH 2779  
 QY 2433 CGTPEPIVNGHNGENYSYRGSVYVQCNAGFRILGMSVRIICODHMSGKTEPCVPI 2492  
 DB 2780 CGTPEPIVNGHNGENYSYRGSVYVQCNAGFRILGMSVRIICODHMSGKTEPCVPI 2839  
 QY 2493 HPGNPVNGLTQNGOFLNVDVYKFCNPGTMAAGARSOCLASQNSDMLPTRIINCTDP 2552  
 DB 2840 HPGNPVNGLTQNGOFLNVDVYKFCNPGTMAAGARSOCLASQNSDMLPTRIINCTDP 2899  
 QY 2553 GHQENSVRQ---VHAGSPRFSFGTTVSRCNHGFLYLTPTLSCQDGTWPRPOCL 2609  
 DB 2900 GHQENSVRQ---VHAGSPRFSFGTTVSRCNHGFLYLTPTLSCQDGTWPRPOCL 2955  
 QY 2610 VSCGHPSPPHSOMSGDSTYVAVVRYSCIGKRLVGNSTRMCGLDGHWGSLPHCSGTS 2669  
 DB 2956 VSCGHPSPPHSOMSGDSTYVAVVRYSCIGKRLVGNSTRMCGLDGHWGSLPHCSGTS 3015  
 QY 2670 VGVCGDPGIPAHGIRLSDGFDGCTMRPSCAGHVLKSSERTCOANGSWSGQPECVI 2729  
 DB 3016 VGVCGDPGIPAHGIRLSDGFDGCTMRPSCAGHVLKSSERTCOANGSWSGQPECVI 3075  
 QY 2730 SCGNPQTSNARVESDGLVFSSSIVYECREGEVYATGLSHRCSVNGTWGSDPECLVIN 2789  
 DB 3076 SCGNPQTSNARVESDGLVFSSSIVYECREGEVYATGLSHRCSVNGTWGSDPECLVIN 3135  
 QY 2790 CGDPGIPANGLBLNDPRYNTKYTCYVGMESHRVSLCTYKDRFNGTKPYCALM 2849  
 DB 3136 CGDPGIPANGLBLNDPRYNTKYTCYVGMESHRVSLCTYKDRFNGTKPYCALM 3195  
 QY 2850 CKPPLIPNGKVGSDPFMWSGVYATACDEGYOLSLPAVFTOENGSGWNGELPOCPVPCG 2909  
 DB 3196 CKPPLIPNGKVGSDPFMWSGVYATACDEGYOLSLPAVFTOENGSGWNGELPOCPVPCG 3255  
 QY 2910 DPGVSRKRREDRGFSYRSVFSFCHPPLVVGSPRRFCQSDGTWSTGTPSCIDPTLTC 2969  
 DB 3256 DPGVSRKRREDRGFSYRSVFSFCHPPLVVGSPRRFCQSDGTWSTGTPSCIDPTLTC 3315  
 QY 2970 ADPGVPOFGIOMNSGQVYVSTVAFRCQKGYLTGSTRCLPMLTNSGTTPDCVPHNCR 3029  
 DB 3316 ADPGVPOFGIOMNSGQVYVSTVAFRCQKGYLTGSTRCLPMLTNSGTTPDCVPHNCR 3375  
 QY 3030 QPETPTANVAGLDPMSGYTLI 3052  
 DB 3376 QPETPTANVAGLDPMSGYTLI 3398

RESULT 5  
 CSM3\_MOUSE STANDARD; PRT; 2796 AA.  
 ID CSM3\_MOUSE  
 AC 080779; Q8BVQ0; Q9D589;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE CUB and Bushi multiple domains protein 3 (Fragment).  
 GN Name=Csm3; Synonyms=Klaa1894;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Brain;  
 RX MEDLINE=22579291; PubMed=12693553;  
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,  
 Nakajima D., Nagase T., Ohara O., Koga H.,  
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:

RT II. The complete nucleotide sequences of 400 mouse K1A-homologous  
RT cDNAs identified by screening of terminal sequences of cDNA clones  
RT randomly sampled from size-fractionated libraries."  
RL DNA Res. 10:35-48(2003).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND SEQUENCE OF 2379-2796 FROM N.A.  
RP (ISOFORM 1).  
RC STAIN-C57BL/6J; TISSUE=Medulla oblongata, and Testis;  
RX MEDLINE=22354683; PubMed=12466651; DOI=10.1038/nature01266;  
RX Okazaki Y., Futuro N., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RX Miyakido I., Otsu N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RX Yagi K., Tomaru Y., Haesgawa Y., Nogami A., Schonbach C., Gotojori T.,  
RX Baidarrelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RX Schirni L.M., Kapapin A., Matsuda H., Baralov S., Beisel K.W.,  
RX Blake U.A., Brad D., Brusic V., Chochla C., Corbani L.E., Cousins S.,  
RX Dalia E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RX Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RX Grummond S., Guslinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RX Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RX Konagaya A., Kurochkin I.V., Lee Y., Lennard B., Lyons P.A.,  
RX Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RX Nagashima T., Nunnata K., Okido T., Pavan W.J., Petrea G., Pesole G.,  
RX Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RX Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RX Sandelin A., Schneider C., Sempke C.A., Setou M., Shinada K.,  
RX Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RX Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RX Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RX Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carinini P., Hayatsu N.,  
RX Hitzoane-Tienhikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
RX Shiraki T., Waki K., Kawai J., Atawa K., Atakawa T., Fukuda S.,  
RX Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RX Miyazaki A., Sakai K., Sasaki D., Shibaek K., Shingawa A.,  
RX Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q80T79-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q80T79-2; Sequence=VSP\_009053, VSP\_009054, VSP\_009055,  
CC VSP\_009056;  
CC Note=No experimental confirmation available;  
CC -1- SIMILARITY: Belongs to the CSND family.  
CC -1- SIMILARITY: Contains at least 10 CUB domains.  
CC -1- SIMILARITY: Contains at least 25 Sush1 (CPC/SCR) domains.  
CC -1- CAUTION: Ref.2 (BAC37116) sequence differs from that shown due to  
CC multiple frameshifts from position 2763.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, AK122567; BAC65849.1; -  
DR EMBL, AK015672; BAB29924.1; -  
DR EMBL, AK078076; BAC37116.1; ALT\_PNAME.  
DR MCD, MGI:1922193; 493050M1ARLK.  
DR InterPro, IPR000859; CUB.  
DR InterPro, IPR00436; Sush1\_SCR\_CCP.  
DR Pfam, PF00431; CUB; 10.  
DR Pfam, PF00084; Sush1; 25.  
DR PROSITE, PS01180; CUB; 10.  
DR PROSITE, PS09223; SUSH1; 25.  
CM Alternative splicing; Repeat; Sush1; Transmembrane.  
CT NON TER 1

|    |           |      |      |  |                            |             |
|----|-----------|------|------|--|----------------------------|-------------|
| FT | DOMAIN    | 1    | 52   |  | Sushi 1.                   |             |
| FT | DOMAIN    | 54   | 162  |  | CUB 1.                     |             |
| FT | DOMAIN    | 167  | 224  |  | Sushi 2.                   |             |
| FT | DOMAIN    | 226  | 336  |  | CUB 2.                     |             |
| FT | DOMAIN    | 339  | 398  |  | Sushi 3.                   |             |
| FT | DOMAIN    | 400  | 508  |  | CUB 3.                     |             |
| FT | DOMAIN    | 511  | 571  |  | Sushi 4.                   |             |
| FT | DOMAIN    | 573  | 682  |  | CUB 4.                     |             |
| FT | DOMAIN    | 685  | 745  |  | Sushi 5.                   |             |
| FT | DOMAIN    | 747  | 855  |  | CUB 5.                     |             |
| FT | DOMAIN    | 928  | 919  |  | Sushi 6.                   |             |
| FT | DOMAIN    | 951  | 1034 |  | CUB 6.                     |             |
| FT | DOMAIN    | 965  | 1026 |  | Sushi 7.                   |             |
| FT | DOMAIN    | 1028 | 1136 |  | CUB 7.                     |             |
| FT | DOMAIN    | 1139 | 1198 |  | Sushi 8.                   |             |
| FT | DOMAIN    | 1200 | 1308 |  | CUB 8.                     |             |
| FT | DOMAIN    | 1311 | 1370 |  | Sushi 9.                   |             |
| FT | DOMAIN    | 1372 | 1483 |  | CUB 9.                     |             |
| FT | DOMAIN    | 1482 | 1543 |  | Sushi 10.                  |             |
| FT | DOMAIN    | 1545 | 1656 |  | CUB 10.                    |             |
| FT | DOMAIN    | 1656 | 1718 |  | Sushi 11.                  |             |
| FT | DOMAIN    | 1719 | 1780 |  | Sushi 12.                  |             |
| FT | DOMAIN    | 1781 | 1845 |  | Sushi 13.                  |             |
| FT | DOMAIN    | 1846 | 1903 |  | Sushi 14.                  |             |
| FT | DOMAIN    | 1904 | 1961 |  | Sushi 15.                  |             |
| FT | DOMAIN    | 1962 | 2019 |  | Sushi 16.                  |             |
| FT | DOMAIN    | 2020 | 2081 |  | Sushi 17.                  |             |
| FT | DOMAIN    | 2082 | 2139 |  | Sushi 18.                  |             |
| FT | DOMAIN    | 2143 | 2200 |  | Sushi 19.                  |             |
| FT | DOMAIN    | 2201 | 2259 |  | Sushi 20.                  |             |
| FT | DOMAIN    | 2260 | 2319 |  | Sushi 21.                  |             |
| FT | DOMAIN    | 2320 | 2377 |  | Sushi 22.                  |             |
| FT | DOMAIN    | 2378 | 2435 |  | Sushi 23.                  |             |
| FT | DOMAIN    | 2439 | 2497 |  | Sushi 24.                  |             |
| FT | DOMAIN    | 2498 | 2557 |  | Sushi 25.                  |             |
| FT | CARBOHYD  | 125  | 125  |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 251  | 251  |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 285  | 285  |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 330  | 330  |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 439  | 439  |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 695  | 695  |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 750  | 750  |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 868  | 868  |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 940  | 940  |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1018 | 1018 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1108 | 1108 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1244 | 1244 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1375 | 1375 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1380 | 1380 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1413 | 1413 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1584 | 1584 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1626 | 1626 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1800 | 1800 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1831 | 1831 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 1951 | 1951 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2021 | 2021 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2041 | 2041 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2188 | 2188 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2247 | 2247 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2256 | 2256 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2283 | 2283 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2297 | 2297 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2307 | 2307 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2365 | 2365 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2453 | 2453 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2611 | 2611 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2618 | 2618 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2701 | 2701 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2707 | 2707 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | CARBOHYD  | 2716 | 2716 |  | N-linked (G1cMn. .)        | (Potential) |
| FT | VAASPPLIC | 1    | 220  |  | Missing (in isoform 2).    |             |
| FT | VAASPPLIC | 221  | 223  |  | TCO -> VAK (in isoform 2). |             |

```

/FILED=VSP_009053.
TCD -> MAK (in isoform 2).

```

```

FT  VARSPLIC 398 409 /FTID=VSP 009054.
FT  AECGASATNNEG -> GTMSAFLICLY (in isoform
FT  2).
FT  VARSPLIC 410 2796 /FTID=VSP 009055.
FT  Missing (in isoform 2).
FT  SEQUENCE 2796 AA; 305500 MW; D62E2FE758F79DD CRC64;
Query Match 53.0%; Score 8997.5; DB 1; Length 2796;
Best Local Similarity 59.0%; Pred. No. 0;
Matches 1534; Conservative 423; Mismatches 564; Indels 77; Gaps 3;

QY 488 NGKRGDLDLGSSISFLCDGEGFLTGQSEITTCYLKESGVVMSAVLRCEAPCGHLTS 517
DB 1 NARRGDNFOLGSSISIVICEEGFIKTQGETITTCILMDGKVMGSPRPGCAPCGGHFA 60
QY 518 PSGTILSGMPGFGYKDALSCAMVLEAORGRIKITTDFRKEVNYDTLEVRDGTYSAPL 577
DB 61 PSGVILSGMPGFGYKDALSCAMVLEAORGRIKITTDFRKEVNYDTLEVRDGTYSAPL 120
QY 578 IGVYHGTQVPOFLISTSNVYLFTSTDKSHSDIGFOLRYERTITLQSDHCLDPGIPVNGOR 637
DB 121 LGSVNGTQVPOFLISSNFILYLTITDNRNNGFKIHESVYVNTICLDPGIPVHGR 180
QY 638 HGNPFYCALVTESCDGYTLSDGEPLCEEPNFQWSBALPSCCALCGGFIQSSGTLSP 697
DB 181 YGHDFSIGSTVSFSCDQGYRLSHREPLCEKNHMSHPLPTCDALCGDVRGPGSTLSP 240
QY 698 GPPDFYNNINCTMIETSHQKGVFTFTHLESBGDYLITENGFTPTPLAQTLGSL 757
DB 241 GYEFYNSINCTVTVTHQKVQFNFTFHLHDHLYLITENGFTPTPLARLTGSEL 300
QY 758 PAPISAGLYGNFTAVQVFFISDFMSYEGFNITSEYDLECEPEVPAVYSIRKGLQGVG 817
DB 301 PSTINAGLYGNFTAVQVFFISDFMSYEGFNITSEYDLECEPEVPAVYSIRKGLQGVG 360
QY 818 DTLFESGPGYRLEGTARITCLGGRRLNNSPLPRCAEAGNSVTGTQGLTSPNFVNY 877
DB 361 DTLFESGPGYRLEGTARITCLGGRRLNNSPLPRCAEAGNSVTGTQGLTSPNFVNY 420
QY 878 NNNHECTYSIOTOPGKGIOLKARAFELSEGDVLYKVDGNNNSARLLGVFSSHMMGVTLN 937
DB 421 ENNHECTYSIOTOPGKGIOLKARAFELSEGDVLYKVDGNNNSARLLGVFSSHMMGVTLN 480
QY 938 STSSSLMLDFTIAENTSKGFEILHSSFEILKCEDPQTPKGYKVADDEGHFAGSSVSFSC 997
DB 481 STSSSLMLDFTIAENTSKGFEILHSSFEILKCEDPQTPKGYKVADDEGHFAGSSVSFSC 540
QY 998 DPYSILRGSEBELLCGERRTWDRPLPTCVABCGGVTVRGEVSGQVLSRGYAPYENLNC 1057
DB 541 NPGYTLHSGSLKCMGTERRAMVPLPSCICACGGKFGKSSGRLISPGYPPYDNLNRC 600
QY 1058 IWTIEAGCTTGLAFLVDETEVHDVLRIMDGPVSGVILKELSGPALPKDLHSTPNV 1117
DB 601 MMWIEVDPGNIVSLQFLAFDEASHDILKVMGSPENEMLLKEVSSGLIPDISHILNIV 660
QY 1118 VLOFSDFTSQGKFIQSVSTATSCNDPGIPQNGSRSGDSEWAGDSTVFOCDPGYALQ 1177
DB 661 TLOFSDFTSQGKFIQSVSTATSCNDPGIPQNGSRSGDSEWAGDSTVFOCDPGYALQ 720
QY 1178 GSAELSCVXENRFQGPSPPTCIAPCGGDLTGPSSVILSPNYPPEPKGCKMDKTVS 1237
DB 721 GGERICIQVENRYFQGPSPPTCIAPCGGDLTGPSSVILSPNYPPEPKGCKMDKTVS 780
QY 1238 PDYVALVNLINLEBGYDLFHLIYDGRDSLPLIGSFYGSQULPGRHIESSNSLFLAFRSD 1297
DB 781 TLYVLSLAFISFISIBENYFLYIYDGPDSNPLIGSFYGSQULPGRHIESSNSLFLAFRSD 840
QY 1298 ASVSNAGFIYDTEMNRESCFDPGSIKXNTRVGSULKIGSSVTTYCHGGEVEGTSTLSC 1357
DB 841 GSVSTYGFLETKAKLRSCFDPGSIKXNTRVGSULKIGSSVTTYCHGGEVEGTSTLSC 900
QY 1358 ILGPDGKPVMMNRPVCTAPCGGQYVGSQVLSPNYPONTSGLICLYFTVTPDYVVF 1417

```

```

DB 901 IMEDDPRPGWNRVLPSCAPCCRSRTGSEGTVLSPNYKXVSDHNCVYSIAVPEKEF 957
QY 1418 GCPAFPHALNDVVEVHDHSGHSRLSLSSGSHGESLPPLATSNQVLLKFAKGLAPAR 1477
DB 958 ----- 957
QY 1478 GFHFVYQAVPRISATQSSVPEPRYGRKLGDSFVGAIYRPECNAGYALQSGPEIECLPV 1537
DB 958 -----AVPRISATQSSVPEPRYGRKLGDSFVGAIYRPECNAGYALQSGPEIECLPV 1010
QY 1538 PGALQOMNSAPPCVPCGGNLTERRGTLISGFEPRYNSLNCVMKTVPEPGAGIQOV 1597
DB 1011 PMSLAQOMNSAPPCVPCGGNLTERRGTLISGFEPRYNSLNCVMKTVPEPGAGIQOV 1070
QY 1598 VSPFTEQWMDSEVFDGADNTVMTLGSFGTVPALMLNSTSNQLYLHFSDISVSAAGPH 1657
DB 1071 VSPFTEQWMDSEVFDGADNTVMTLGSFGTVPALMLNSTSNQLYLHFSDISVSAAGPH 1130
QY 1658 LEYKTVGLSSCEBPVAVNSGVTKERYLVNDVVSFQCEPGYALQGAHISCMPGTVRRMN 1717
DB 1131 LEYKTVGLSSCEBPVAVNSGVTKERYLVNDVVSFQCEPGYALQGAHISCMPGTVRRMN 1190
QY 1718 YPPPLCIAQCGTVEEMGVILSPGFPNGYSDNMDCKMILPVGFGAHIQELNSTEPN 1777
DB 1191 YPPPLCIAQCGTVEEMGVILSPGFPNGYSDNMDCKMILPVGFGAHIQELNSTEPN 1250
QY 1778 HDYIIRNGPYETSHMWRFGSGELPSLLSTSHETTVYFPHSDHSONRPGKLEYOAYEL 1837
DB 1251 HDYIIRNGPYETSHMWRFGSGELPSLLSTSHETTVYFPHSDHSONRPGKLEYOAYEL 1310
QY 1838 OECPPPEPFANGIYRGAANYNGQSTPECLPGYOLTGHPVLTQCHGTRKANDHLPKCEV 1897
DB 1311 OECPPPEPFANGIYRGAANYNGQSTPECLPGYOLTGHPVLTQCHGTRKANDHLPKCEV 1370
QY 1898 PCGGNITSNNGTVSPGFPSPYSSSODCVMLITVPIGHVLANLTLQTESGSGFITWMD 1957
DB 1371 PCGGNITSNNGTVSPGFPSPYSSSODCVMLITVPIGHVLANLTLQTESGSGFITWMD 1430
QY 1958 GPQCTAPPLGVFTSMKAKTVYQSSSNQVLLKFRHDAATGCIIFALFASAYPLTKCPPTIL 2017
DB 1431 GPQCTAPPLGVFTSMKAKTVYQSSSNQVLLKFRHDAATGCIIFALFASAYPLTKCPPTIL 1490
QY 2018 PNAEVTENEEFNIGDLYRYRCLEPGFTLVNGEILTCGLGYTLQEGSPPTCEVNCPTNEL 2077
DB 1491 PNAEVTENEEFNIGDLYRYRCLEPGFTLVNGEILTCGLGYTLQEGSPPTCEVNCPTNEL 1550
QY 2078 LTTSTGVILSGSYPGFQFQTSQWLVREPDYNISLTVETFLSEKQYDFEFLPDGSGQ 2137
DB 1551 LTTSTGVILSGSYPGFQFQTSQWLVREPDYNISLTVETFLSEKQYDFEFLPDGSGQ 1610
QY 2138 SPILKALSGNYSAPLITSSNSNYLWMSDHAANKRGFKIRISAPYCSIPRAPLHGFIL 2197
DB 1611 SPILKALSGNYSAPLITSSNSNYLWMSDHAANKRGFKIRISAPYCSIPRAPLHGFIL 1670
QY 2198 GQTSSTOPGSGIHFGCNAGYRLVGHSMALCTRRHOGYHLEMEALPLCOALSQGLPEAPXG 2257
DB 1671 GQTSSTOPGSGIHFGCNAGYRLVGHSMALCTRRHOGYHLEMEALPLCOALSQGLPEAPXG 1730
QY 2258 MWEKKEYTGTKAVYSCSEGYHLQAGAETAECLDTGLMSNRNVPQCVVTCPDVSSIS 2317
DB 1731 MWEKKEYTGTKAVYSCSEGYHLQAGAETAECLDTGLMSNRNVPQCVVTCPDVSSIS 1790
QY 2318 VEHGRMLIFETQYQOQAOQLMLCDGGYVYTGQVLRRCQANGKMSGLDSTPTRTIISGGE 2377
DB 1791 VEHGRMLIFETQYQOQAOQLMLCDGGYVYTGQVLRRCQANGKMSGLDSTPTRTIISGGE 1850
QY 2378 LPIEPNGHRIGTLVYGATLIFSCNAGYTLVGSRYVECMANGGLMSGEYRCLAGHGCTPE 2437
DB 1851 LPIEPNGHRIGTLVYGATLIFSCNAGYTLVGSRYVECMANGGLMSGEYRCLAGHGCTPE 1910
QY 2438 PLYNGHINGENYYSRQSVYQCNAGRLLIGMSVRLICQDHHMSGKTPFCVPTTCGHPNP 2497

```

```

Db 1911 LIVNGQVIGENVYEDTVVYOCNPGFRLIGSSVRIQQDHNWSQLPSCVPEVSCGHPSP 1970
Qy 2498 VNGLTGQNGFNLNDVVKFCNPGVYAGABASQCLASQSDMLPTCIIINCTRGHEN 2557
Db 1971 IYGTSGGFGFNNDVVTSCNIGYLMGPTKAOCCANQWMPVPVCKVAVNCSPGIPAN 2030
Qy 2558 SVRQ---VHAAGPHRFSGFTTVSYRCNHFYLLGTVPVLSGQDGTDRPRQCLLVSCGH 2614
Db 2031 SKRESKIEHGN-----FTYGTVPYDCNPGYFLPSSSVLICQPNQMDPELECIIMDCGH 2086
Qy 2615 FGSPPHSGSDSYTVGAIVNVRSCIGKXTLVGNSTRMCGLDGHTGSLPHSCGTSVGVCG 2674
Db 2087 PGIPNVALSGEKNYFCSTVHYVSCGKRSLLGQASRTCOLNGHNSGSPHSCGDTGTGCG 2146
Qy 2675 DPGIPAHGIRLGDSEDFCTVWRFSCEAGHVLRGSEERTCOANGMSGQPEPCGVISGCPN 2734
Db 2147 DPGIPGHSRQSDPRTKSTVRFACDGTIYLIGSEERTCLNSGWTGKQPECKAVQCGNP 2206
Qy 2735 GTPSNARVVFSDGLVFSSSIYVECEGEGYATGLSRHCSVNGTWGSDPECLVINCGDPG 2794
Db 2207 GTTANGKVFRIIDGTFSSSVIYSCLEGYILSGPSVROCTANGTWSGSLPNCITIIISCDPG 2266
Qy 2795 IPANGRLIGNDFRYNKTVTYOCVPGYMHSHRVSVLCTKDRYVNGTKPVCKALMCKPP 2854
Db 2267 IPANGRLYGDVFGVQNVSYMCQPGYTELNGSRVRCCTTNGTWSGVVPTCRAVYCTSTP 2326
Qy 2855 LIPNGKVVGSDPMGSSVTVACLEGYOLSLPAVFTCEBNGSWTELPCCFVFGCDPVP 2914
Db 2327 QISNGRLEGTFNPDGFSISYICSAAGYELSPAVLITCVNGTWSGVPCLKFCGDPPIR 2386
Qy 2915 SRGRREDGFSYRSSVSFSCHPPLVYSGPRRFGSDGTWGTGTOPSCIDPTLTTCADPGV 2974
Db 2287 SQGRREGSFTYQSEVSFSCSPFLVSSSTRLCOTDGTWSSGSHCIEPRTISCENGV 2446
Qy 2975 PQFSIQQNSQGYQVGSYVLPFCOKGYLLQGSTTRTCLNPLTWSGTPDPVPHKQPEPTP 3034
Db 2447 PRHSGQNNTPFGFYQVGSVVOFHKCKGHLQGSTTRTCLNPLTWSGIQPECIHSCQPEPSP 2506
Qy 3035 THAVGALDLPMSGYTLI 3052
Db 2507 AHAHVGMGLDPSHGTYLI 2524

```

## RESULT 6

```

ID 072241 PRELIMINARY; PRT: 1466 AA.
AC 072241;
DT 01-JUN-2003 (TREMBLREL. 24, Created)
DT 01-JUN-2003 (TREMBLREL. 24, Last sequence update)
DT 01-MAR-2004 (TREMBLREL. 26, Last annotation update)
DE SI: b21C3.1 (Novel protein similar to CUB and Bush1 multiple domains
  protein (CSDM)) (Fragment).
GN Name=SI:b21C3.1;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
  Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Babbage A.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL732489; CAD60796.1; -.
DR HSBP; Q9UCV4; INZ1.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00431; CUB; 9.
DR Pfam; PF00084; Sush1; 8.
DR SMART; SM00032; CCP; 8.
DR SMART; SM00042; CUB; 9.
DR PROSITE; PS01180; CUB; 9.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; UNKNOWN_1.

```

```

DR PROSITE; PS00923; SUSH1; 8.
FT NON_TER 1
FT TER 1466
SO SEQUENCE 1466 AA; 160456 MW; 45F45ED456FF472E CRC64;

Query Match 36.4%; Score 6180.5; DB 2; Length 1466;
Best Local Similarity 74.1%; Pred. No. 0;
Matches 111; Conservative 170; Mismatches 181; Indels 37; Gaps 4;

135 MDAHLNPGSGITSPNPPIQYDNNANCWITLNLNPSKYIKALFEEPDLERGYDTLTVG 194
2 MCGQLGPGSGILTSPNFPVQYDNNANCWITLNLNPSKYIKALFEEFDELERGYDTLTVG 61
Qy 195 DGGDDGDKTYLVYQNAQSDSPHTPSRIPE---SNSGDIWRKMTVLEICRISSDA 251
Db 62 DGSYIGDKRTVFNH-----LSGTTTDLVYSTSHQWMLN----- 95
Qy 252 RSGSVRSKPSITSNABELVAPGTEIEQSCGCGPGIPAYGREGSRFHHDGTLKPECOPAF 311
Db 96 -----FKTDDTSSGLGFKVGEIEIDQGCCDPPGIPAYGKREGIFRRGDKLYFECLEPAF 150
Qy 312 LVGQXAITCOKNQWSAKKQCVFSCFPNFTSPSGVLSVNYPEPDYGNHLHCWLLIARP 371
Db 151 LVGKKNITCOKNQWSAKKSCVFSCEFNFTTSPSGVLSVNYPEYGNHHCWLLIARP 210
Qy 372 ESRHILAFNDIVPEPDPFLVIKDGATAEAPVLTGTFSGNQLPSITSSGHVAREPQTDH 431
Db 211 ESRINLAFNDLSMEKQDPFLSTIKDGAESPILTSTFGSDVLPSTITSGHVARLEPLTDH 270
Qy 432 STGKRGFNITFTFRHNECEDPPGVVNVKGFSGDLSQSSISFLCDEGFLTQSEETITC 491
Db 271 TYDGRGNITFTFRHNECEDPPGVVNVKGFSGDLSQSSISFLCDEGFLTQSEETITC 330
Qy 492 VLKGSVAVNSAVIRCAPCGGHLTSPSGITLSPGMPFYKDALSCAMVIAOQYPIKI 551
Db 331 ILKQGNVAVNSAVIRCAPCGGHLTSPSGITLSPGMPFYKDALSCAMVIAOQYPIKI 390
Qy 552 TFDKFTENVYDTELVVDGRTYSAPLIGVHGVPOVPELITSTSNVYLSTDSKSHDIG 611
Db 391 IFDKFTENVYDTELVVDGRTYSAPLIGVHGVPOVPELITSTSNVYLSTDSKSHDIG 450
Qy 612 FOLRYETITLQSDHCHDPCGIPVNGQRHGNFVYALVTFSCDSGYTLSDSEPLECEBNFQ 671
Db 451 FRIRYETITLQSDHCHDPCGIPVNGQRHGNFVYALVTFSCDSGYTLSDSEPLECEBNFQ 510
Qy 672 WSRALPSCDALCGGFIQSSGTLISPGFPDPFYNNLNCWIIETSHKGVYFTFTFHL 731
Db 511 WSRALPSCDALCGGFIQSSGTLISPGFPDPFYNNLNCWIIETSHKGVYFTFTFHL 570
Qy 732 SGHLYLITENGSTPTQLRDLTGSRLPAPISAGLYGFTQVRFISPSMSYSEFNITFS 791
Db 571 SPHDHLVITENGSTPTQLRDLTGSRLPAPISAGLYGFTQVRFISPSMSYSEFNITFS 630
Qy 792 EYDLEPCEBEVPAVYSIRKGLQFVGDTLTFSCPPGRSLRGSTARITCLGRRRLMSPLP 851
Db 631 EYDLEPCEBEVPAVYSIRKGLQFVGDTLTFSCPPGRSLRGSTARITCLGRRRLMSPLP 690
Qy 852 RCVAECNSVYTGQGLTLLSNFPVNNVNHCEIYISIQTPQKGIQLKARAFELSEGDVLK 911
Db 691 RCVAECNSVYTGQGLTLLSNFPVNNVNHCEIYISIQTPQKGIQLKARAFELSEGDVLK 750
Qy 912 VYDGNNSAALLGVFSHSEMAGTLLNSTSSLMIDFTDLENNSKGEFLHPSSEELIKCE 971
Db 751 VYDGNNSAALLGVFSHSEMAGTLLNSTSSLMIDFTDLENNSKGEFLHPSSEELIKCE 810
Qy 972 DPGPKFGYKVDHGGHAGSSVSFSCPGYSGLGSEELLCLSGERRRTWDRPLPCVAECG 1031
Db 811 DPGPKFGYKVDHGGHAGSSVSFSCPGYSGLGSEELLCLSGERRRTWDRPLPCVAECG 870
Qy 1032 GYVAGEVSGQVLSPGYPAPYEHNLNCIWTIEAGAGCTIGLHPLVFTDEVHVDVLRWDGP 1091
Db 871 GKIKDESMGRLISPGYPAPYEHNLNCIWTIEAGAGCTIGLHPLVFTDEVHVDVLRWDGP 930

```

```

QY 1092 VESGVLLKELSGPALPKDLHSTFNSVLIQSTDFSTKSGFAIOPFVSTATSCNDPGI RQ 1151
D 931 QGGVTLRELSGSTLPDLHSTFNSVLIQSTDFSTKSGFAIOPFVSTATSCNDPGI RQ 987
QY 1152 NSRSGDSWEADSVFQCDPBGALQSGAEISCVKIENRFFWQSPPTCIAPCGGDLTGP 1211
D 988 NSTRIGDSREPDHVLFOCDPGYLLQAGAKITCTEINNRFWQSPPTCIAPCGGDLTGP 1047
QY 1212 SGVILSPNYPPEYPPGKCDMKVTVSPVIALVINFLNPGVDFHLYNGRSLPLI 1271
D 1048 SGLILSPPEYPPGKCDMKVTVSPVIALVINFLNPGVDFHLYNGRSLPLI 1107
QY 1272 GSFYSGQLPGRTESSNSLFLAFRSDASVSNAGFVIDYENRSCDPCGIRKNGTVGS 1331
D 1108 GSFYGTDPVDRTESSNSLFLAFRSDASVSNAGFVIDYENRSCDPCGIRKNGTVGS 1167
QY 1332 DLKLGSSVTVYCHGVEYEGTSTLSCILGPDGKPPVNNRPRVCTAPCGQYVSGDYVLS 1391
D 1168 ELKLGATVTVYCDSTYLGSDATLTCIMGSDGKPKGKPKVCIAPCGGQYVSGDYVLS 1227
QY 1392 PNYPONTYSGQICLYFVTVPKDYVVFQGFAPFHTALNDVVEVHDGSHSLSLSGSH 1451
D 1228 PGYPNVYSSGRTCLYSVVPRDYVVFQGFAPFHTALNDVVEVHDGSHSLSLSGSH 1287
QY 1452 TGESELPATSNQVLIKPSAKGLAPARGHFHYQAVPRTSATQSSVPEPRYKRLGSDFS 1511
D 1388 TGESELPATSNQVLIKPSAKGLAPARGHFHYQAVPRTSATQSSVPEPRYKRLGSDFS 1347
QY 1512 VGAIVRECEGNSGVALQSGPEIECPVPGALAQWNVASPTCVPCGGVLTERRGTTLSRPG 1571
D 1348 VGAVVSEFCNAGVYLEBPSAIECLTVRNALAQWNVASPTCVPCGGVLTERRGTTLSRPG 1407
QY 1572 PEPILANSICVWKIVVEGAGIQIQQVVSFVTEQNMDSLEVFADGADNTVLMGFSFGTV 1630
D 1408 PEPILANSICVWKIVVEGAGIQIQQVVSFVTEQNMDSLEVFADGADNTVLMGFSFGTV 1466

```

## RESULT 7

Q9ES77 PRELIMINARY; PRT; 3567 AA.

```

AC Q9ES77;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Polymorphism precursor.
GN Name=Polymorphism.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH/Hensic;
RX MEDLINE=20517255; PubMed=11062057;
RA Gilgus D., Vint M.-A., Callebaut I., Coulombel L., Cacheux V.,
RA Romeo P.-H., Vigon I.;
RT "Polymorphism: a secreted protein with pentraxin complement control
RT protein, epidermal growth factor and von Willebrand factor A
RT domain";
RU Biochem. J. 352:49-59(2000).
DR EMBL; AF206329; AAG32160.1; -.
DR HSSP; Q9UCV4; INZI.
DR MGD; MGI:192849; Polymorphism.
DR GO; GO:000509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR009030; Growth_fac.
DR InterPro; IPR003410; Hyal_in.
DR InterPro; IPR001759; Pentaxin.

```

```

DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00008; EGF_7.
DR Pfam; PF07645; EGF_CA_2.
DR Pfam; PF02494; Hyr_2.
DR Pfam; PF00354; Pentaxin; 1.
DR Pfam; PF00084; Sushi; 33.
DR Pfam; PF00092; VMA; 1.
DR PRINTS; PR00010; EGFBLDOD.
DR PRINTS; PR00895; PENTAXIN.
DR PRINTS; PR00453; VWFADOMAIN.
DR ProDom; PD002153; Pentaxin; 1.
DR SMART; SM00032; CCP; 34.
DR SMART; SM00179; EGF_CA; 6.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 6.
DR PROSITE; PS00022; EGF_1; 9.
DR PROSITE; PS01186; EGF_2; 11.
DR PROSITE; PS00026; EGF_3; 9.
DR PROSITE; PS01187; EGF_CA; 6.
DR PROSITE; PS00825; Hyr; 2.
DR PROSITE; PS00923; SUSHI; 34.
DR PROSITE; PS0234; VMPA; 1.
DR EGF-like domain; Signal.
FT SIGNAL 1
SQ SEQUENCE 3567 AA; 387393 MW; 8FBA8276E12293E5 CRC64;

```

Query Match 11.0%; Score 1864.5; DB 2; Length 3567;  
 Best Local Similarity 21.4%; Pred. No. 3e-97;  
 Matches 793; Conservative 376; Mismatches 1173; Indels 1359; Gaps 148;

```

QY 99 TCNEGYDQSGKRIKCMKVSMDPFAWSDHRCARMDALRGRSGITTPNPIQYDN 158
D 466 TCNEGYDQSGKRIKCMKVSMDPFAWSDHRCARMDALRGRSGITTPNPIQYDN 512
QY 159 NAHCWIIITLNPSSKVITKLAFFEEFDLERGYDTLTVGCGQDGDQKTVLYMQNACSDSPH 218
D 513 NAHCWIIITLNPSSKVITKLAFFEEFDLERGYDTLTVGCGQDGDQKTVLYMQNACSDSPH 520
QY 219 TPGSRIPSMGSDIMRQKTVLEICRDISSDASGSRKSPKTSNAVELVAPTEIEQG 278
D 521 TPGSRIPSMGSDIMRQKTVLEICRDISSDASGSRKSPKTSNAVELVAPTEIEQG 567
QY 279 SCGPBGIPAYGRARRGS-----RHHGDTLKFCOPAFE-----LVGQKAITGCKN 323
D 568 SCGPBGIPAYGRARRGS-----RHHGDTLKFCOPAFE-----LVGQKAITGCKN 627
QY 324 NQMAKAKKPGCVFS-----CFNFTSPSGV-VLSPNYPEDYGNHLHCWVILA 369
D 628 NQMAKAKKPGCVFS-----CFNFTSPSGV-VLSPNYPEDYGNHLHCWVILA 671
QY 370 RPESRHIAFNIDIVERPODFPVLVIXOGATAEAPVLGTFG-GNOLPSSITSSGHVAREFQ 428
D 672 RPESRHIAFNIDIVERPODFPVLVIXOGATAEAPVLGTFG-GNOLPSSITSSGHVAREFQ 706
QY 429 TDHSTGRGNNITFTFRHNECDPVGPNVNGKRFSGSLQSGSISFLCDGFGTQGS-E 487
D 707 TDHSTGRGNNITFTFRHNECDPVGPNVNGKRFSGSLQSGSISFLCDGFGTQGS-E 765
QY 488 TITCVLKEG-----SVVWNSAVLRCEAPCG-----GHUTSPSGTI 522
D 766 TITCVLKEG-----SVVWNSAVLRCEAPCG-----GHUTSPSGTI 825
QY 523 LSPGMPGFYKDA-----LSCAWVIEAQPGY-----IKITPRRFT 558
D 826 LSPGMPGFYKDA-----LSCAWVIEAQPGY-----IKITPRRFT 885
QY 559 EVNYDTLEVVDGR-----TYSAPLIGVHGT-----QVPOFLIST- 593
D 886 EVNYDTLEVVDGR-----TYSAPLIGVHGT-----QVPOFLIST- 945
QY 594 ---SNLYLLFSDKSHSDIGPOLARETITLQSDHC-----LDPGIPVNGQR----- 637

```

Db 946 ETTIRRLKLNKEMYS---FQLASERVADNSLITEKAFLECRPGSVLRGMVCNCP 1002  
 Qy 638 HGNDFYVALVTFECDSGYTLSDGEPLCE-----PFWMSRALPSCCLCGGFIQGS 690  
 Db 1003 LGTISYLEHSTCECLMSGYDEBQLECKLCPPRTHTTEYLHRSVSECKAQCKOGTYSS 1062  
 Qy 691 SG-----TI-----LSPG 698  
 Db 1063 SGLTECSCLPTGYOPERFSRSCLLCPETTTTVKRAVDISACGVPCPVSEFSGLTFC 1122  
 Qy 699 FP---DFYPNN-----LNCWT-----IETSHG 718  
 Db 1123 YPCRDYQYQNAKGSFCLACPFYGTITTTGATSIYDCSSFSSTSAEESIIVAPGHS 1182  
 Qy 719 KGVE---FTPHTHL-----ESGHDYLLITENGSTOPLRQITG-----754  
 Db 1183 QNKKEVSSQVHEGECFLNCHNSGTCCQQLGRGYVCLCPRG-----YTGKCEITDDEC 1234  
 Qy 755 SRLPAPISAGL---YGNFTAQVRF-----ISDFMSY---783  
 Db 1235 SSLPC-LANGICROOVGFTCECSLGYSQICBENINCEISSPLCNKGTCTDGLASTHCT 1293  
 Qy 784 --EGFNITFSEYDELPCEPEVPAYSIRKGLQFGVGDITLTFSC---FPGYRLEGTA--834  
 Db 1294 CVKGMGVHCEETDVNECQSSPLNNAVCXKQVGG---FSCCKCPRGFLGTRECKNVD 1347  
 Qy 835 -----RITCLGGRRLNSSPLRCVABEGNSVTGTGTLNBNFVNPN-----878  
 Db 1348 CLSOPCONGATCKKQANSF-----RC--QCPAGFTGHCEP-----NINECOSNPCR 1392  
 Qy 879 NNHECI-----YSIOTOPKGIQLKARAFELSEGDVLKVYDGNNSARLLGVFHSHEMG 933  
 Db 1393 NQATCVDELNYSCKCQPG---FSGHRCETBP-----SGFLNDEVSIGYGVVLDD 1442  
 Qy 934 V-TLNSTSSLMW--DFIT-----DAENTS---KGFELHFSFELI---968  
 Db 1443 VLPILHAITCAFMKMSDVINYGPISYALBDDKMTSLTLDYNGWVLYVNGKEKITNCP 1502  
 Qy 969 -----KCEBPGT-----975  
 Db 1503 SVNDGIWHHIAITWTSTGAMRVYINGELSDGGTGLSIGKAIPEGGALVLQEBODKKEG 1562  
 Qy 976 -----PKF-----GYKVHD 984  
 Db 1563 FNPABSFVGSISQNLMDYVLSPOOVKLLASSCEBELSRGNVLAHPDFLSGITKVYDS 1622  
 Qy 985 EGHFA-----GSSVSPSCDPGYSLRGSBELCLSGERRTDRP 1022  
 Db 1623 SSMFCSDPSEGSVPHLRPASGNRKPGSKVSLFCDPFGVMGNPVQYCLN--QGOWTQF 1680  
 Qy 1023 LPTCV-AEGC--GTARGEVSQVLSPGYRPAHYENHNLNCTITAEACCTIGLHFL---1074  
 Db 1681 LPHCEIRICGRLPRLAENGVSABEDFHAAGSTVYQ-----CTSGYVILGDSR 1726  
 Qy 1075 VFDEEYVADVLRINDGPVSEGVLLKELS-----1102  
 Db 1727 MFCIDN-----GSMNGISPSCLVDDECAVSDCSEHASCNTNNGSYVSCNPPYTGDSKN 1781  
 Qy 1103 -----GPALPKDLSTFNSVVLQFSTDEFTSKOGFALQFS-----VSTAT--1142  
 Db 1782 CABPVKCAKAPENFENGHS-----SGEITYT--VGTAVTFPSDEGHELVGVTITCLE 1830  
 Qy 1143 -----SCNDPGIIPONGSRSGDSWEACDSIVFOCDPGYALQSAEISCVKIE 1188  
 Db 1831 TGEWDRLRPSCEALISCGVPVPVENGVDGSAFTYGSKVYVRCDDGYTLISGEBESACIASG 1890  
 Qy 1189 NRFFMQSPPTC-IAPCGGDLTGPBGVILSPNYPEPYPRGECOMKWTVPDVIYA----1243  
 Db 1891 S---WSHSPVCELVKC-----SQBEDINNGK-----YILSGITY 1922  
 Qy 1244 LVFNIENLEPGYDLHAYIDGRDLSPLIGSFYGSQLPGRISSSNSLFLARSDASVNA 1303  
 Db 1923 LSIASYSCENGYSL-----QGFELLECTASGSH---DRAPPS 1957

Qy 1304 GFVIDYTENDRESCEPDGSIKNGTRVGSDDLKASSVYTYCHGGEVEGTSTLSCLAPDG 1363  
 Db 1958 QLV-----SGEPPYKDAVTTSSNTTFGTAVYTKEGEGLTGPDTIVC--QANG 2006  
 Qy 1364 KPVNNRNPVCTAPCGQYVGSQVLSPPNYPONTYSGOICLVFVTPKD--YVFGQFAF 1422  
 Db 2007 K--WNSNHQCLA-----VSCBEPVVDHAS-----PETAHRLFGDTAF 2043  
 Qy 1423 FHTLANDVNEVHDGSHSLLSLSGSHTGESLPLATSNQVLKISAKLAPRGHFV 1482  
 Db 2044 YYCA-----DQYS-----LADNSQLICNAQGNVPPAG-----2071  
 Qy 1483 YQAVPRTSAPQCSVPEPRYG---KRLGDSFVGAIVRFCONSGYALQGSPEIECLVPVG 1539  
 Db 2072 -QAVPRCIAHCEKRPVSYSILESVAKAPAGSVSPKMBEGLNNTSKIECLR--2127  
 Qy 1540 ALAQMNVSAPTCVPCGANTLERGTILSPGFPPYLSINCWKIIVPBGAGIQIOVS 1599  
 Db 2128 -----GEMS-----PSPL--SVQC-----2140  
 Qy 1600 FVTEQNMDSLEVPDADNTVTLGFSFGTVPALLNSTSNQVLHYFYSDISVSAAGHLE 1659  
 Db 2141 -----IPV-----2143  
 Qy 1660 YKTVGLSSCEPRANVPSNGVTERGYLVNDVVSFOCEBEGYALQGAHISCMPGTVRMNYP 1719  
 Db 2144 -----RCBPPSIANGYPSGTNYSFGAAVAASHKGFYTKKSKSTC--BATGWSKP 2194  
 Qy 1720 PPLICIAOCGGTVEBEGVILSPGPGNYPNMDCSWKIALPVGGAHIOFLNFTSEBNHD 1779  
 Db 2195 TPTC-----HP--VSCNEBPXYENGLTHTTGRTEFSEARFQ 2229  
 Qy 1780 YIEIRNGPYETSRMNGRFPSSSELPSLSLTSHTTTFPHSDHSONRPFLEIYOAYELO 1839  
 Db 2230 C---NPGYXA-----AGS--FVFVCOANR---HMHSADAPLSTPL-----N 2262  
 Qy 1840 CPDPEBANGIVGAGVNGQSVTFECLPQYOLGHHVULCOHNTNNMD--HPLPKCEVP 1898  
 Db 2263 CGKRPLONGFLKESFEVGSKVOFCNBEGLVGDNSWTCQ--KSGKMSKRPK--VP 2319  
 Qy 1899 --CGGNTSSNGTVYSGFPSPYSSQDCWMLTVPIGHGVRLNLSLQTEPSGDFPTIW 1956  
 Db 2320 TKCAEPPLLENQVLK-----LASEVGMTISCKEGHALQ--GSPVLKCLPSSQ-----W 2368  
 Qy 1957 DGPOQTAIRLGVTFRMAKKTVOSSNOVLKTHRAATGATPAIAFSAVPLTK--CP 2013  
 Db 2369 NG-----SFPICGMVLCP 2382  
 Qy 2014 PTLIPNAEVTTENEERFIDIVRYRCLPGFTLVGNELITCKLGYLOFBEPPPTC--EVHC 2072  
 Db 2383 PPLIPFG-VPASSGALHFGSTVKYLCVDFGLRSGPILICQADS--TWSSPLPFCVBEVC 2439  
 Qy 2073 PTNELTDSGTVLISQSPGYSYPOFOTCSMLVREPPYNSLTVEYFLSEKQYDEFEIFD 2132  
 Db 2440 PQBEIILN--GIHVGOLALSTLLYTC-----KPFEL-----VGNATTLIC 2479  
 Qy 2133 GPSQO-----SPLLKALSNGYSABRLITSSNSVYLWSSDHAYNRKGFKIRIYAPYCSLP 2188  
 Db 2480 GENQOMLGKPMCKPIB-----CPEP 2500  
 Qy 2189 RARLHGFILQOTSPQPGSIIHFGCNAGYLVGHSMAICTHPOGYHMSAIPICQALSC 2248  
 Db 2501 KEILNG--QFSVSPYOYQOTITTFCDRGRFLRLEGPKSLTCLBTGD--WMDPBCDAIHC 2555  
 Qy 2249 GLPEAPKNGVAVGKEVYVGTAKAVYSCSEGNLOAGAATAECLDTGLMSNRNVPPOCVPV 2308  
 Db 2556 SDPQRIENGVBADYVYGAAMIYISCPRGQVLGHAMQT--CEBSG--WSSSS--PTCVPI 2610  
 Qy 2309 TC---PDV---SSISVEHGRMLIF-----T 2329  
 Db 2611 DCGLRPHIDFGDCTKVVDGQHFDQEDMDMEVYLAHPQHLIATAKALENTKESPAHAS 2670

```

QY 2330 QYFOAQLMLICDPGYVYTGVRVIRCOANGKSLDSTPTCRISCGELPI-PPNGHRIG 2388
DB 2671 HFLYGMVSYSCPEPGLLIGIPVLLIQDEGTWN--GTAPSCISIC-DLPAEPNGFIHF 2727
QY 2389 TLISVGATATPSCNSGYTLVGSVRVECANGLMSSEVRCLAGHGTREPIVNGHNGEN 2448
DB 2728 TQTYWGAQAQYSCKHGHIIEGSHLRLCLQNKQMSGTVPRECAISCSKNPILMNS IKDD 2787
QY 2449 YSRGSSVYQACNAGFRLLIGMSVRIQQODHNSGKTPFCVPTITCGHPGPNVGLQGNQFN 2508
DB 2788 YSLDVLVYEDDSGYTLINGSKKRTQENRDMGDHEPMCLPYDCCSPVPTVIRGVKGEYT 2847
QY 2509 LNDVVKFCNPGVMAAGASQCLASGQMSDMLPTCRILINCTDGHQENSRYQVHASGPH 2568
DB 2848 FQKEITVSCREBFIIEGARSRLCLTNQSGWSGATSCMVRCPAPQPVNGV----AGLD 2903
QY 2569 RPSFCTVSYCNHGFYTLGTIVLSQCGDGTWDPRPCCLLVSCGHGSPHSGMSGDSY 2628
DB 2904 -YGFKEVAFHCLBSGYVLOGAPRLTQSGNGTWDAEVPCKPATCGPPADLPQGFPGFSF 2962
QY 2629 TVGAVVRYSCIGKRTLVGNSTRMCLDGHWTGSLPHC-----SGTSVGV- 2672
DB 2963 YHGHGIQYQCFPTGYKLNPNRRLCLPNQSGSSPCLPCRCSTPTIIOGTINATIDGCG 3022
QY 2673 -----CGD-PGIPIAHGIRLGDSPDPTV 2694
DB 3023 KTVQIECFKPKLLGLSEITCDANGQSDVPLCEHAQCGPLPTIP-NAIVLEGLSEBDV 3081
QY 2695 MRFCSEAGHVLRGSSERTCOANGSMGSOPECGVISCNGTPTNA----- 2740
DB 3082 VTYSGRPGVTWQSSDLICTKALMSQPIYCEPLSCGPPTVANAATGEAHTYESKVK 3141
QY 2741 -----RVVF-SDGLVFSSSIYV 2756
DB 3142 LRLCEGVMSDITDTFTQODGHWVPERITCSPKCKVPSNMTRIRHGDPOVNGQSV 3201
QY 2757 ECRGYATGLSLHNCVNGTWTC--SDPECLVINCDDPGIPANGRLGNDFFYNKTVY 2814
DB 3202 SCABGFHIEGVNWSCTCPDGPDPFSDSCIPVCHPSPSAHSGVVGKHSFGSTIYV 3261
QY 2815 QCVPGYMMESHRYVLSCTKDRITWNGTKPYCKALMKCPRLPNKVYSGDFFMGSSVYV 2874
DB 3262 QCDPGYVLEGNREI--COENRQMSGEVAVCRNRCETPAEPFNGRAVLENTTSGSLIF 3319
QY 2875 ACLEGYDLSIPAYFTCCGNSWTELFQCPVFCGDPGV-PSRRREDRDFSYRSSVSFS 2933
DB 3320 SCHRGYTLBSPEAHCTANGTWNHLTPCLCKNRCVPVPIPENAVLSKEFYVDQVSIK 3379
QY 2934 CHPLVLVGSPPRFCCGSDGTWSTGTOPSCIDPILITCADPGVPFGIQQNNSG--YVGS 2991
DB 3380 CREGFLLKNGVITCSDETHWTNARC--EKISCGPSHVENAL--ARGVYQYQDM 3433
QY 2992 VLFRCQGYLLQGSTTRTCLPNLTWSTGTPRDCVPHHCROPEPTHTANVALDLPS----- 3046
DB 3434 ITYSYCSGYMLEGSLRSVCLENGTM--TPSPVCAVCRP-----CQNGVCCQAPNACSP 3487
QY 3047 ---MGVTLILPARRASPRVAPSTAPARRMAQAARPSAW 3084
DB 3488 DGMGRILC-----EPPICILPLANGRCVAPYQCDCCPTGM 3522

```

```

RESULT 8
016744 PRELIMINARY; PRT; 2489 AA.
AC 016744;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Complement receptor 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_Taxid=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94065175; PubMed=8245463;
RA Vuk D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RL and sequence of the coding region unique to the S allele."
RN [2]
RP SEQUENCE FROM N.A.
RA Vuk D.P., Wong W.W.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: L17428; AAB60695.1; JOINED.
DR EMBL: L17429; AAB60695.1; JOINED.
DR EMBL: L17418; AAB60695.1; JOINED.
DR EMBL: L17430; AAB60695.1; JOINED.
DR EMBL: L17392; AAB60695.1; JOINED.
DR EMBL: L17394; AAB60695.1; JOINED.
DR EMBL: L17396; AAB60695.1; JOINED.
DR EMBL: L17398; AAB60695.1; JOINED.
DR EMBL: L17400; AAB60695.1; JOINED.
DR EMBL: L17409; AAB60695.1; JOINED.
DR EMBL: L17408; AAB60695.1; JOINED.
DR EMBL: L17407; AAB60695.1; JOINED.
DR EMBL: L17406; AAB60695.1; JOINED.
DR EMBL: L17405; AAB60695.1; JOINED.
DR EMBL: L17403; AAB60695.1; JOINED.
DR EMBL: L17402; AAB60695.1; JOINED.
DR EMBL: L17401; AAB60695.1; JOINED.
DR EMBL: L17419; AAB60695.1; JOINED.
DR EMBL: L17417; AAB60695.1; JOINED.
DR EMBL: L17416; AAB60695.1; JOINED.
DR EMBL: L17415; AAB60695.1; JOINED.
DR EMBL: L17414; AAB60695.1; JOINED.
DR EMBL: L17413; AAB60695.1; JOINED.
DR EMBL: L17412; AAB60695.1; JOINED.
DR EMBL: L17411; AAB60695.1; JOINED.
DR EMBL: L17420; AAB60695.1; JOINED.
DR EMBL: L17427; AAB60695.1; JOINED.
DR EMBL: L17426; AAB60695.1; JOINED.
DR EMBL: L17425; AAB60695.1; JOINED.
DR EMBL: L17424; AAB60695.1; JOINED.
DR EMBL: L17423; AAB60695.1; JOINED.
DR EMBL: L17422; AAB60695.1; JOINED.
DR EMBL: L17421; AAB60695.1; JOINED.
DR EMBL: L17420; AAB60695.1; JOINED.
DR EMBL: L17399; AAB60695.1; JOINED.
DR EMBL: L17397; AAB60695.1; JOINED.
DR EMBL: L17395; AAB60695.1; JOINED.
DR EMBL: L17393; AAB60695.1; JOINED.
DR EMBL: L17391; AAB60695.1; JOINED.
DR EMBL: L17390; AAB60695.1; JOINED.
DR PIR: I73012; I73012.
DR HSSP: P17827; IGRN.
DR GO: GO:0004872; F:receptor activity; IEA.
DR InterPro: IPR000834; Peptidase M14.
DR InterPro: IPR001424; SOD CU ZN.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF000084; Sushi_37.
DR SMART: SM00032; CCR_37.
DR PROSITE: PS00133; CARBOXYEPT_ZN_2; UNKNOWN_3.
DR PROSITE: PS00087; SOD CU ZN_1; UNKNOWN_1.
DR PROSITE: PS50923; SUSHT_37.
SQ SEQUENCE 2489 AA; 272846 MW; CEI1B53F2B4FA6 CRC64;

```

```

Query Match 9.1%; Score 1550; DB 2; Length 2489;
Best Local Similarity 20.8%; Pred. No. 1.9e-79;
Matches 676; Conservative 335; Mismatches 1031; Indels 1204; Gaps 141;

```

```

Qy 117 VSDMFAASDHRPYCRARMCDAHLRGPSGITSPNFQYDNNACWIIIT-ALNPSKV 175
Db 147 ISGOTVIMDNETPICDRIPC-----GLPPTITNGDFTSTRENHYSVATYRCNPSGG 201
Qy 176 KLAEEFPELBERGYDITLVGDGODGODQKTVLYMSGNACSDSPHTPGSHIPESMGDIWRQ 235
Db 202 RKVEE-----LVGEPS-----LYCTSN-----DDQV-IV-- 225
Qy 236 KMYULEICRODISSDARSGSVKSPKTSNAVELVAPGTEIQQSGCDP---GIPAYGR 291
Db 226 -----SGPAPQCIIIP-----NKCTPVENEGILVSDNR 253
Qy 292 EGRFPHGDTLKFEQAPAFELVGOKAITCOKONOMSAKPGCVFSCFNFSPSGVILSP 351
Db 254 --SLFSLNEVVEFRQCPGFVKGPFRVYCOALNKKEPELPSCSRVC-----QPPDVLHAE 307
Qy 352 NYPDYGNHLHCWMLILARPESRIHLAFNDIDVEPODFLYIKDQATAEAVLGTFSGNQ 411
Db 308 RTQBDKNF-----SPQGEVFS-----CEGYD-----LRGAASMRCTPGQDMS--- 347
Qy 412 LPSITSSGHVAREFOTDHTGKRGFNITTTFRHNECPD--PGVPVNGK-RFGDSIQG 469
Db 348 -PAA-----PTCEVKSDDPFMQLNGRVLFPVNLQIG 379
Qy 470 SSISFLDEGFLTQSGSETITCVLKEGSVWNSAVLRCEAPCGHILSPSGTILSPGMG 529
Db 380 AKVPFVCEGFR-QLKSSASVYCVLAGMESLWSSPVCE-----QIFCPSPPVLPNG-- 430
Qy 530 FYKDALSCAWIEAOPGYPIKITEPDRFTEVNYD-----TLEVR 568
Db 431 -----RHTGKELEV--PFGKAVNYTCDDPHDRGTSDLIGESTIRCTSPQ 475
Qy 569 DGRYYS--APLIGYHGTQVPOFLISTSNLYLFSTDKXSD--IGQLAYE----- 617
Db 476 GNGVWSSPAPRCGLIGHCOAP-----DHFLFAKLKTQTNASDPPIGTSLEYCRPEYGG 529
Qy 618 ---TITLOSHCLD-----PGIPVNGGRG--NDPVGALVMTFSCDSG 655
Db 530 RPSFIT---CLDNLVWSSPDKVCKRKSCTPPDPVGMVHTVDIOVGRINYSCTG 584
Qy 656 YTLSDGEPLCEBPN---FQMSRALPSCBAL--CGEFTQSGSGTILSPGFPDEYNNLNTW 711
Db 585 HRLIGHSSAECLISGNAHMTKPIQORIPG----- 617
Qy 712 IITSHGKGVFTPTHTFLHESGHDYLLITENGSTQPLRLQTLGSRLEPAPISAGLYXNETA 771
Db 618 -----LPTIANG----- 625
Qy 772 QVRPIISDF-SMSYSGFNITFSEYDLPECEPEVPAVYSIRKGLQGVGDTLTFSCFPG-- 827
Db 626 ---DITSTNRENFH-----YGSVATYRCNPPSGGG 651
Qy 828 ---YRLBETARITCIGERR--LWSSPLPRCAEGCNSVTGTGTLSPN--FPVNYNN 880
Db 652 RKVELVGEPSIYCTSNDDQVIGMSGAPQCI-----XPNCTPNNENG 696
Qy 881 HECIYSIQTOPKGIQLKARAFELSEGDLKVYDGNNSARLLGVFSHSEMAGVTLNSTS 940
Db 697 -----ILVSDNRS-----LFSINBY-- 711
Qy 941 SSLWLDFTTDAENTSKEPELHFSSEFELIKCEDPGTPKPKGYKHVBGHRAGSSVSFSCDPG 1000
Db 712 -----VEFRQCPG 719
Qy 1001 YSLGSEELCLGSEBRTMDRLPTCAVECGTVAGEVSGVLSPGYAPAEHNLNICTWT 1060
Db 720 FVMGCPRRVKCOALNK--WEPELPSCSRVC-----QPP-- 751
Qy 1061 IEAEAGCTIGLHFLVFTBEVADVLRIMWDGVEVSGVLLKELSGPALPKDLHSTNSVVLQ 1120
Db 752 -----DVLH----- 755

```

---

```

Qy 1121 FSTDFTSKOGFALIQFESVSTATSCNDPGLPONGSRGSDWEAGDSTVPOCDPGYALOGSA 1180
Db 756 -----AERTORDDINSPPQGEVYVSCBPGDILGAA 786
Qy 1181 EISCVKIENRFPWOPSPPTIAPCGDLTPR--SGVILSPNYPEPYPGKCBMKWTVPSP 1238
Db 787 SMRCTPOGD--WSPAFTCEVKSDDPFMQLNGRVLFP----- 823
Qy 1239 DYVALVFNIFNIEPGDPLFIHYDGRSLPLIGSFYGSQLPGRIBESSNSLFLA----- 1293
Db 824 -----VNIQLGAKVDVCD-----EGFQDKG--SSASTCVLAGMESL 858
Qy 1294 FRSDASYNAGFVVDYTENPRESCFDPGSIKNGTRVSDKL--GSSVYTYG-- 1344
Db 859 WNSSVVPCEQIF-----CPSPPVLPNGHCTKPLEVFPFGTIVYTCDPHDRG 907
Qy 1345 GGYEBEGTSTLSCILGPDGKPVWNNPPPVCTAPCGQVGSDDVILSPNYPQNTSQIC 1404
Db 908 TSFPLIGESTIRCTSDPDQNGVWSSPAPRC-----GIL-- 940
Qy 1405 LYFVTYKDYVVGQFAFHHTALNDVENVHDSQ--HSRLLSLSGSHTGESLPLATSN 1462
Db 941 -----GHCQADHFLFAKLKTQTNASDPPIGTS- 968
Qy 1463 QVLIKFSAGLAPARGF-----HFVYQAVPR--TSATQSSVPEPRYK--RLGSDPSVG 1513
Db 969 ---LKYECREBYRGPRPSTICLDNLVWSS--PKDYCKRKSCTPPDPVNGMVHTVDIOVG 1024
Qy 1514 AIVAFECNSGVALQGSPEIECLPVPGALQMNVSAPTGV--VPCGNLTERGTLSPGFP 1572
Db 1025 SRINVSCTGHLRILGHSSACI--LSGNAHMTKPIQORIPGLPPTIANGDFTSTNRE 1083
Qy 1573 EPIYLSNLCWAKIVPEAGAGIQIOVYFVTEQUMDSLEVEDGADNTYTMIGSFGTIVPA 1632
Db 1084 NFHYGS-----VATYRCNPPSGGRKVE-----LVGEPS-- 1112
Qy 1633 LNSTSNQYLHFSDISVBAAGFHELYKTVGSSCEBPVPSNGVKTGER--YLVNDV 1689
Db 1113 -IYCTSNDDQVIGMSGAP-----QCIXPKCTPPNV--ENGILVSDNRSLSFINBY 1161
Qy 1690 VSPQCEPGVALQGHANISCPGTVRRWNPYPPLICIAOCGGTVEBMEGVILSPGFPNYS 1749
Db 1162 VEFRCQGFWMKGRBRVKQ--ALNKWEPELPSCSRVC----- 1197
Qy 1750 NMDCSKIALPVGRGAHIQFLNPSTEPNHYIETERNPETSMMGRFSGSELPSLSLT 1809
Db 1198 -----QPPDVL-- 1204
Qy 1810 SHETTYFHSQONRPGFLEYQAYELQCPDEPPANGIVRGAGVNGQSYTFECTLPG 1869
Db 1205 -----HAERTOR-----DKDNF-----SPQGEVYVSCBPG 1229
Qy 1870 YOLTGHPVLTCOHTNENMDHPLPKCEVPCGAGNTSS--NGTVYSPGFPSPSSQDCW 1927
Db 1230 YDLRGASMTCT--POGDWGPAPAPTCEVKSDDPFMQLNGRVLFP----- 1273
Qy 1928 LITVPIGHVRLNLSLQTEBPGDFTIIMGPOQTARLGVFTRSMAKKTVOSSNOVL 1987
Db 1274 -VNLOLGAKV-----DFVCD--EGFQDKG--SSASTCVLAGMESL 1302
Qy 1988 KFHMDATGFIAPAFSAVPLTK--CPPTIILPNAEVTEN--EEFNIGDIYVRCLP-- 2041
Db 1303 -----AGMESLWNSVFPCEQIFCPSPPVLPNGHCTKPLEVFPFGAVNTYCDPHR 1354
Qy 2042 ---GFTLVGNBILTKL-----GTYLQFEGPPICEV--HCPTNELLTSTGVLISQSY 2090
Db 1355 DRGTSFLLIGESTIRCTSDPDQNGV--SSAPARCGIILGHCOA----- 1395
Qy 2091 PGSTPQOTQCSWLVRWEPDYNISLTVEYFLSEKQYDFEFLFDGSGQSPILKALSGYSA 2150
Db 1396 -----PD-----HFLFAKLKTQTNASDPPIGTSLEYCRPEYGR 1430
Qy 2151 PLIVTSSNSGVLRWSS--DHAVNRKGFKIXYSAVYCSLPAPALHGFILGQTSNQPGSGIH 2209

```

```

Db      1431 PFSITCLDNLV---WSSPKDVCCKRS-----CKTPPDVPMGMVHVIDIVGSSIN 1478
Qy      2210 FGCMNGYRLVGHSMALCTRHPOGYHLMSEALPLCALSCGPEAPKXNMVG---KEYTV 2266
Db      1479 YSCTGHHLIGHSSAECLISGNTHA-WSTKPPICRIICGLPPTIANGDPISTRENFHY 1537
Qy      2267 GTRAVYSCSEGYH---LQAGAEATAECL---DTGLMS-----NRVVP----- 2303
Db      1538 GSVVYTRCNLGRKRVKVELVEPSIYCTSDNDQVIMSGARPCIIIPNKCTPNNWNGI 1557
Qy      2304 -----OCVP-----VTC-----PDVSSIS-----VEHGR 2322
Db      1598 LVSDNRSLFSINEVEFRCCQPFVWKPRRVKCALNKEBELPSCSRVCCPPEILHGE 1657
Qy      2323 WRLIFETOYFOAOLMLICDPGYYYTQGVRIQCQANGKWSIGDSTPICRISIC---GEL 2378
Db      1658 HTPSHQDNFSPGOEVFYSCBPGYDIRGASLHCTPQGDWS--PAPRCVAVKSCDPLGOL 1715
Qy      2379 PIPNGHRIGTSLV-YGATAIFSCNSGYTLVGSVRBECMANG--LMSGSEVRCLAGHC 2434
Db      1716 ---PHGRVLPPLNLQLGAKVSVFVCDDEGRLKSSVSHCVLGMSLNNVVPCEHIFCP 1772
Qy      2435 TPEDIVNGHNGE---NYSYRGSVYVQCN-----AGRLIGMSVRIICQDDH---WSG 2481
Db      1773 NPAILNGRHTGTPSGDI PYGKEISYTCDPHPRGMFTNLIGESTICTSDPHGNGWSS 1832
Qy      2482 KTRPC-VPTTCG-----PGNPVNGLTQGNQFNLDVYKFCFNGVMAEGARS 2529
Db      1833 PAPRCISLVARGHCKTPEQFPFASPTIPINDF---EPVGTSLINECRGRTGKMFSTIS 1888
Qy      2530 QCLASGOWMLPTCRILINCTDPGHQNSVRQVHASGPHRFSGTIVSYRCHNFPYLLGT 2589
Db      1889 -CLENLWSSVEDNCRKRS CGPPEPENGWNIHTD---TQGSTIVNCSNGFRLIG 1943
Qy      2590 PVLSCQGDG---TWDRRPOCLVSCGHPSPHSGMSGDSYR-----VGAVVYXSC- 2638
Db      1944 PSTTCLVSGNNTVYDKAPICEIISIC---EPPELTISNGFYNNRTSPHNGVTVYVYQCH 1999
Qy      2639 -----IGKRLVGNSTRMCGLDGHWTSGLPHCSGTSYVCGDPIPAHGRILGD 2687
Db      2000 TGPDEQLFELVGRSIY--CTSKMDQGVWSSPPRCISTN--KCTAPBV-ENALRVG 2054
Qy      2688 S---FDGTYMFRSCCEGHVLRSSERTCANGSWSSGSOPECVIGCGNPTPSNARVVF 2744
Db      2055 NRSFSLTEIVRRCCQGFVWVGSHTVQCTNRMGPXLPHCGRV--COPPELHGEHTL 2113
Qy      2745 S--DGLVFSSSIYVECEGYATGLLSRHCSVNGTWTGSDPECLVINCGD--PGIPANGL 2800
Db      2114 SHQDNFSPGOEVFYSCBPSYDLRGAASLHCTPOGDMSPAPRCTVKS CDFLQOLPHGRV 2173
Qy      2801 RLGNDRYKNTVYVOCVPGYMWESHVVS-VLSCTDRKTNGTKPVCKALMKCPPLPIPN 2858
Db      2174 LLEPLNLQLGAKVSVFVCDDEGRLKGRSASHCVLAGMK-ALMNSVVPCEQIFCNPALIN 2232
Qy      2859 GKTVGS---DFWNGSSVTVAC-----LEGQLSLPAVFTCEGNGSTGELPQC- 2903
Db      2233 GRMTGTPRFDIPYGKEISYACDTHPRGMFTNLIG--ESSIRCTSDRQNGVWSSPAPRCE 2291
Qy      2904 --FVPECQDPGVRSRGRREDRGFSY--RSSVFSFCHPLVLVGSFRRRCQSDGTWSTGQ 2958
Db      2292 LSVPAACDPBPRLKQNHGYIGHNSLPLPMTISYIDPGLVLGKGFICTPOGILWSQID 2351
Qy      2959 PSCIDTTLTTCADPGVFGIQLNN--SGCYOVGSFVLRCKGKYLQGSTRTTCLPILNT 3015
Db      2352 HYCKE---VNCSPF-LFANGISKELEMKVHYGVYVTLKCBDYTLGSSPWSQCADDR 2407
Qy      3016 WSGTTP 3021
Db      2408 WD--PP 2411

```

```

07LCS3
ID 07LCS3 PRELIMINARY; PRT, 3494 AA.
AC 07LCS3;
DT 05-JUL-2004 (TReMBLrel, 27, Created)
DT 05-JUL-2004 (TReMBLrel, 27, Last sequence update)
DE 05-JUL-2004 (TReMBLrel, 27, Last annotation update)
DE Intrinsic factor-vitamin B12 receptor (Fragment).
GN Name-CUBN;
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SOURCE FROM N.A.
RA Aminoff M., Brady S., Verroust P.J., Moestrup S.K., Krahe R.;
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 4 EGF-like domains.
DR EMBL; AF243129; AAK61830.1; -.
DR EMBL; AF243083; AAK61830.1; JOINED.
DR EMBL; AF243084; AAK61830.1; JOINED.
DR EMBL; AF243085; AAK61830.1; JOINED.
DR EMBL; AF243086; AAK61830.1; JOINED.
DR EMBL; AF243087; AAK61830.1; JOINED.
DR EMBL; AF243088; AAK61830.1; JOINED.
DR EMBL; AF243089; AAK61830.1; JOINED.
DR EMBL; AF243090; AAK61830.1; JOINED.
DR EMBL; AF243091; AAK61830.1; JOINED.
DR EMBL; AF243092; AAK61830.1; JOINED.
DR EMBL; AF243093; AAK61830.1; JOINED.
DR EMBL; AF243094; AAK61830.1; JOINED.
DR EMBL; AF243095; AAK61830.1; JOINED.
DR EMBL; AF243096; AAK61830.1; JOINED.
DR EMBL; AF243097; AAK61830.1; JOINED.
DR EMBL; AF243098; AAK61830.1; JOINED.
DR EMBL; AF243099; AAK61830.1; JOINED.
DR EMBL; AF243100; AAK61830.1; JOINED.
DR EMBL; AF243101; AAK61830.1; JOINED.
DR EMBL; AF243102; AAK61830.1; JOINED.
DR EMBL; AF243103; AAK61830.1; JOINED.
DR EMBL; AF243104; AAK61830.1; JOINED.
DR EMBL; AF243105; AAK61830.1; JOINED.
DR EMBL; AF243106; AAK61830.1; JOINED.
DR EMBL; AF243107; AAK61830.1; JOINED.
DR EMBL; AF243108; AAK61830.1; JOINED.
DR EMBL; AF243109; AAK61830.1; JOINED.
DR EMBL; AF243110; AAK61830.1; JOINED.
DR EMBL; AF243111; AAK61830.1; JOINED.
DR EMBL; AF243112; AAK61830.1; JOINED.
DR EMBL; AF243113; AAK61830.1; JOINED.
DR EMBL; AF243114; AAK61830.1; JOINED.
DR EMBL; AF243115; AAK61830.1; JOINED.
DR EMBL; AF243116; AAK61830.1; JOINED.
DR EMBL; AF243117; AAK61830.1; JOINED.
DR EMBL; AF243118; AAK61830.1; JOINED.
DR EMBL; AF243119; AAK61830.1; JOINED.
DR EMBL; AF243120; AAK61830.1; JOINED.
DR EMBL; AF243121; AAK61830.1; JOINED.
DR EMBL; AF243122; AAK61830.1; JOINED.
DR EMBL; AF243123; AAK61830.1; JOINED.
DR EMBL; AF243124; AAK61830.1; JOINED.
DR EMBL; AF243125; AAK61830.1; JOINED.
DR EMBL; AF243126; AAK61830.1; JOINED.
DR EMBL; AF243127; AAK61830.1; JOINED.
DR EMBL; AF243128; AAK61830.1; JOINED.
DR GO:0004872; F:receptor activity. IEA.
DR InterPro; IPR000152; Aex_hydroxyl_3.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR009030; Grow_fac.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00431; CUB; 27.

```

DR Pfam; PF00008; EGF, 4.  
 DR Pfam; PF07645; EGF CA; 3.  
 DR SMART; SM00042; CUB; 27.  
 DR SMART; SM00181; EGF; 8.  
 DR SMART; SM00179; EGF CA; 7.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 4.  
 DR PROSITE; PS01180; CUB; 27.  
 DR PROSITE; PS00022; EGF\_1; 4.  
 DR PROSITE; PS01166; EGF\_2; 2.  
 DR PROSITE; PS00026; EGF\_3; 5.  
 DR PROSITE; PS01187; EGF CA; 3.  
 DR EGF-like domain; Receptor.  
 FT NON TER 1  
 SEQUENCE 3494 AA; 384281 MW; 04A811043985DB39 CRC64;

Query Match 8.78; Score 1473; DB 2; Length 3494;  
 Best Local Similarity 21.28; Pred. No. 8e-75;  
 Matches 795; Conservative 416; Mismatches 1184; Indels 1356; Gaps 174;

113 TCMKVSMPAA-----WSDHRCVRA--RMCDALHRCPSGIITSPNPIQVNNACVMI 165  
 317 TCVDGVDSFSCETRLMTG--ALCVPOQVCBSLSGINSFSYRSPDVGVHVNCFWV 374  
 166 ITALNPKVKIKLAEEFDLER---GYDTLVGDGQDQKTVLYMSQNAQSDSPHTPG 221  
 375 IKT-EMKVLKRTFTFRLESMDCRHFLOY---YDGSASAFOGRFGSSLPH--- 426  
 222 SRIPESMGDIWROKMTVLEICRDISSDARSQSVKRSPTKSNALVAPGTEIEGSCG 281  
 427 -----ELLSSD-----NALVFLYSEHLRNG--- 447  
 282 DPGI PAYGRRGSRFHHDITLFCORAFELVQKAITCOKNMWSAKKPCVNSCFPNF 341  
 448 -----RGFTVR-----WETQOP---ECGIL 465  
 342 TSPGVVLSPNYPEDYGNHLHCWLLIARPEIRIHLAFNDIDVPR---QFDPLVIKGA 397  
 466 TGPVGIKSPGYPNYPGRDCWIVVTSPLVLTFTFGTISLEHDDCKMDVLEIDGP 525  
 398 TAEPVLTGSPGNOLPSSITSSGHVAREFQTDHSTKRGFNITFTFRIN-ECPPDGPV 456  
 526 LYODPRLGKCTTSPVPLQTTGPFAIHFDSDQISDQGFHIYLTSPSLRCGNYTD 585  
 457 VNGRRF-----GDSLOG-SSISFLCD-----EGF 480  
 586 PBEGLFPELISGPFTTRQCVMMKOPQGEQIQINFTHVELQCSDSQNYIEVRDETL 645  
 481 LG-TOGSEITTCVLKEGSVWV-----NSAVLRG--EAPCGHLTSPSGTILSPGP 528  
 646 LGKVCNGGTISHIKSTNSVWIRKIDASVEKASFRAYQVACGDELTG-EGVIRSPFP 704  
 529 GFYKDALSCAMVIEAOPGYPIKITFDRF---TEVNYDTELVNDRGTYSAPLIGVYGT 584  
 705 NVYPERGTCRMTIHQPOSQVILNFTVEIGSSAHCEITDVEIGSSIIILSPENKXCGT 764  
 585 QVPEPLISTNMYLYLFTSDKSHSDIGFQLYETITLQSDHCDLPGI PVANGQRGNDFV 644  
 765 DIBPFTSYVNFVLTIVTKSSSTENHGFMAK----- 796  
 645 GALTTFSCDSDGYTLSDGEPLCEBPNFQMSRALPSCAL-CCGFIOGSSGTILSPGPDFY 703  
 797 -----SAEDLACGEILTESTGITIISGSHNVY 823  
 704 PNNLNTCAIITSHGKGVFTFHTFLE---SGHDYLLITENGSTFOPLRQLTGSRLPA 759  
 824 PHGINCTMHLIVQPNHLIHLMFETFLHFNCTNDYLEVYDTSETS-LGRYCGKSIIP 882  
 760 PISAGLYGNFTAQVRFISDPSMSYEGNITFSEVLDLAPCEBPVPAISIRKGLQPGVGT 819  
 883 SLTIS--GN-SLMVLPVYDSDLAEGFLINEAI----- 913  
 820 LTFSCPGYRLGCTARITCLGRRRLWSSPLPRCAECGNSVTGTGTLSPNPNVYNN 879

914 -----SAATACL-----QDYTDLCTFTSPNPNVYNN 941  
 880 NHECTISIQTPKGIOLKARAFELSRG-----DVLKATVDGNNAKRLIGVSHSEMG 933  
 942 NMECTIYRTVQQLAVHFTNFSLEAIGNYYTDFLEIRDGGEKPELGIIF-YGSNLP 1000  
 934 VTINSTSSLMPLFITDAEYTSKGFELHFSFELIKEDGTPKFGKYVHDEGHFAGSSV 993  
 1001 PTIISHNKLMLFKSKQIDITRSGFSAYWD-----GSSTCGKLLTSSG 1045  
 994 SFSQDPCY-----SLRSEELL-----CLSGERRTDRP--- 1022  
 1046 TF-ISPVPMPVYHSSRCYMWLSSHGSAFELEKDFLHHNPCTLDLYAVYDGPSSNS 1104  
 1023 -----LPTCAECGTV-----RG-----EVSQVLS 1044  
 1105 HLLTOLGDEKCPRLIRSSGDSMFILKRTDGOGRGKAEYROTCEENVIVNQTGILES 1164  
 1045 PGYAPAEHNLNCMTTEAGCTIGLHFLVPTEEVH-----DVLRIWDGVSGLLK 1099  
 1165 IGTPNPISEKHQCMWITRAITGNTVYTLAFDLH-HINCSTDYELVYDGPOMG--- 1219  
 1100 ELSGPALPKDLHSTFNSVVLQFSTDFPTSKQ-GFALQPSV-----STAT-SCNDGP 1150  
 1220 RYGVVDLPFGSTTSSKQLVLLTDGVRGKRGFQWQMPVYGGGELSGATGFSFSPGF 1279  
 1151 QNSRSGDS-WEAGDSTVPCDPCYALQ-----GSAEISCVKI 1187  
 1280 NRYPNKCEIWM-----YIRTDGSSIQITIHDFVEYHRCNFDVLEIYGGDPFHSRI 1333  
 1188 -----EN-----RF-----FMOPSPPTCIAPCGGDLTGPS 1212  
 1334 AOLCTORSPEPMQVSTGNEALIRFKTDLISNGRCNASWQ---AVTGGCGGIFAPPS 1389  
 1213 GVILSPNYPPEYRPGKECDWKVTVSPDYVALVFNIENLPEGYDFLHIYDGRDS-LSPLI 1271  
 1390 GEIHSPPVPSPYRSTNDSCAVIVRYRHYRLVLPFDLEDPQSCIMAYDGLSTWGRLA 1449  
 1272 GSFTYSGQLPGRISSSNSLFLAFRSDASVSNAGFVIDYTNPRESC-----FD--- 1319  
 1450 RTGREGALNPIVSSGNSLFLRFOSGFSRQNRFRQF---RQACGHLITSSFDVSS 1505  
 1320 ---PGSIKNGTR---VGSDLKGSVTVYCHGVEYEGSTLSC---IL--GPDGKPV 1366  
 1506 PREPAPNPNQNSWIIQAPPLNHTLSFTH--FELERSTICARDEVELLDGHEAPL 1563  
 1367 -----WNNRPVCT-----APCGQYVGSQGVVLS 1392  
 1564 RGRYCGTDMHPHTTSPSALTLRFVSDSSISAGFHTTVYASVACGCTFYMAEGIFNSP 1623  
 1393 NYQONTYSGQICLY-FVTVPKDYVVFQGFAPFHTAL-----NDVVEVHDSHQSRLL 1444  
 1624 GYDPIYPNVEICWNNIISSPGNRL---QLSFISFOLEDSODCSRDFEIREGNA-----T 1675  
 1445 SSLSGHSGESLPLATSNQV---LKFSAKGLAPARGFFVYQAV----- 1486  
 1676 GHLVGRYCGNSFPLNTSSIVGHTLWRFISDSSGSGTGFOATPMKIFGNDNIYTGKVA 1735  
 1487 -----PRTSATQ-----CSSVPEPRY----- 1502  
 1736 SPFWPENYPHNSNYQMTVVNNASHVHGRILEMDIEIQNCYDXKLRIYDGPRIHARLIG 1795  
 1503 ---GKRLGDSFYGAIVFECNSGVALQGSPEIECLVPGALAOW-NVSAPTCVVP--- 1554  
 1796 AYCQTOTESFSSGNSLTFHFYSDSSISGR-----GFLLEWFAVDAPDVGLPTIAP 1846  
 1555 ---GKRLGTERGR-----IISGRPEPYLNLNCVMKIVPEGAGIQOVVSPTEQON--- 1605  
 1847 GACGGFL--RTGAPVFLFSPGMPDYSNNVDCITMLIQADST-VELNITSLDIESHRTC 1903  
 1606 -WDSLEVFADADVTVMGLSFGSTVAPALINSTNOLYLHFYSDISVASAGFLHYKTVG 1664  
 1904 AYDSLVIRDDNNLQAOGLAVLCGRREIRGPIRSIGEVYFIRFTSDSVTRAGFNASFH--- 1960

QY 1665 LSSCEPAVNSGVKTBGRYLNDVVSFOC-----EBCYALQCH-----AHISCM 1709  
 DB 1961 -KSCGGYVLAHADRGIITSPKPIETYSNLNCSWHVLVQSLTAVHEOPFOIPNGDSSCN 2019  
 QY 1710 PG---TVRRMN---YPPPL-----CIAQ----- 1726  
 DB 2020 QBDYVLNRGPDYIPPLGPPGNGHFGGSHASSTLFTSDNMVQPIISDHSNCGCFKI 2079  
 QY 1727 -----CGGTV-----EEMGVILSPGPGNYPSNNMDCSKIALPVGFAHIOF---LN 1771  
 DB 2080 KYEAKSLACGGNNVYHDAISAGYVTSPPHPPHYPHADCIWILAPETRIQLQFEDRFD 2139  
 QY 1772 FSTEBN---HDYLEINRGEYETSRRMGRFSGSELSLSTSHETTYVPHSHQNRGFK 1829  
 DB 2140 IEVTNCTSNYLELADGDVSDAPILSKCTGSLSSQMSGCEWVWYLFRRSDNSFTHVGF 2199  
 QY 1830 LEQYAYELQECDEPPFPFANGIYRGAGY-----NVGOSVTECLPGYOLT----- 1873  
 DB 2200 AK---YSIAOCGRVPQSGVYESIGHPTLPYRDLNLFCEMHLQGLSHYLLTISFEDFNQ 2256  
 QY 1874 ---GHPVLTQCHGNRMND-----HPLP----- 1893  
 DB 2257 NSSG-----CEKDVVEIWDNHTSGNIGRYCGNTIPDSIDTSNTAVRREVTDGSVYASG 2311  
 QY 1894 -----KCEVPQGNITSSNGTVSPGFPSPSSODCWLITVPIGHGYRL---NLSLQ 1945  
 DB 2312 FRLNFESSMEBCGDLQOSIGTFSPYNNPHGRICEMRITAPBRRIITLMEFNRLA 2371  
 QY 1946 TEBS---GDFITIMGPOOTAPRLGVFTRSM-AKKTVOSSSNQVILKFRDAAT---GGIIPA 2000  
 DB 2372 THPCNNHEVYVFNIGINSFPQLEKCSAVVNSNEIKSGNTMKVJFEFTGSRPYGG-FT 2430  
 QY 2001 IAFSAVPLTKCPPTLIPNAEVVTENEEFNIGDIVRCLPGFTLVGN----- 2048  
 DB 2431 ASYSSSDAVCCGS--LPN---TPBGNFT---SPGYDGRVNSRNINCEWTLS 2475  
 QY 2049 -----ELTCKLGYLQFEGP-----PRICE 2069  
 DB 2476 NPNQGNSSISIHEDFYLESODCOQFVLEFRVY---DADGPIMLRCLGPKSKPTLVLVP 2532  
 QY 2070 ---VHCPTELL-----TD-----STGVILISQSPSSYPOFQTCSTLVAV 2106  
 DB 2533 YSQWVHFVITNERVEHIGFAKYSFTDCGGIQTIGDSGVTISPPYAPAYSLTHCSMLTFA 2592  
 QY 2107 EDPYINISLVEYFLSEKO---YDEFEIPDPSGSGPLLKALSGNYSAPLIVTSSNSYV 2162  
 DB 2593 POGHTITLTFSDPDIEPHTTCAMDSTVNRNGSPESPITIGQYCGN-SNRTTIOSSNOLV 2651  
 QY 2163 LRASSDHAKYRKGFKIRISAPYCSLPRAPLHGFILOQTSTOPEGSIHFGCNAGYRLVGH 2222  
 DB 2652 VTNSHSLQGGP---YATWNTQT---LQCGG---IFHS 2682  
 QY 2223 MAICTHPQGYHLMSEAIPLCALSS-CGLPEAPKN-GWFEKBYTVGTAKAVYSCSEGYHL 2280  
 DB 2683 DNGTISPH---WPNPENSRCSTWALTTHSKHEIFDNNFLIPSDGOCQCSFYKV 2738  
 QY 2281 QAGAEKTAEC-LDTGILMSNRNVPPOCVPTCPDVSSISVEHGRWMLIFETQYQFOAOLML 2339  
 DB 2739 WAGTEVEVDKALLATGC---GNVAPG---PVITPSNFTAV-----FQSGRA- 2778  
 QY 2340 ICDPGYVYTGQRYIRCOANGKMSLG-----DSPTCRITISGCELPPLPPNGHRIG 2388  
 DB 2779 ---PAGFSASFVSRGSGNFTGPGYITISPNYPKOYDNNMNTYV-----LEAN----- 2824  
 QY 2389 TLSVYGATAI-FSCNSGYTLVGRVRECMANGLMGSEVRCLAGH-----CG--T 2435  
 DB 2825 PLSVVLLTFVSPHLEARSALTGS---CVNDG-----VHTRKGSVWSTPPTATYCGDEM 2874  
 QY 2436 PEPI-VNGHINENYYSRGSVYVQCNAGRRLIGSVIRICQODHMSGKTPCVPTTCGP 2494  
 DB 2875 PAPTLAGVLLNLFYSNEQITDF---GPK---FSYRI-----ISCGGV 2911

QY 2495 GNPVNGELTQGNQPNLNDVVKVCNPGY-MAEGAASQCLASQWSDMLPTCRINCTDPC 2553  
 DB 2912 FNFSSGI-----ITSPAYADVPNDMHCLYITTVSD----- 2943  
 QY 2554 HOENSVROVHAQPHRFSGTIVYSTRCNHGFLLGTVPVLSQCGDGTWDRPPOCLLVSCG 2613  
 DB 2944 ---DKVIELKFS-----DEDVVPSTSCSHDYAL-----YOGANTSD---PLGKFCG 2985  
 QY 2614 HPGSPPHSQMSG-----DSYTVGAVVRVSCIGKRTLVNSTRMGSLDHMGKSLPH 2664  
 DB 2986 SK-RPENVASNNMSMLVFKTDSFOYAKGKMSF-RQTL-GPOGCCG-GYLTG----- 3034  
 QY 2665 CSGTSVGVGDGCIPAHGIKRLDSDPCG-----TVNRFSCEAGHYLR- 2706  
 DB 3035 ---SNTPASPDSDSNGW---YDKMLNCWIIIPVKNKVIHLTFMTFALEASTQR 3085  
 QY 2707 -----GSSEPTCOANGSMSSQOBCGYISGNGCTGPNANVVSDDLVESSSIY 2756  
 DB 3086 CLDYVYKLYDQSEANANLA-GTFCSGTVPAPFISGNEFLT-----VQFISDLTL 3133  
 QY 2757 ECREGYAT-GLLSRHC---SVNGTWTG---SDPECLVINCGDPGIPANGRLGNDFRYNK 2810  
 DB 3134 E-REGFNATYTIMDVPCCGTVAWTQNISSP-----NSSDDPVFS----- 3175  
 QY 2811 TVTYQCVGYMMESHRSVLSCTKDRTWNGTKPVCKALMCKRPPLIPNGKVVGSDFMWS 2870  
 DB 3176 -----ICTWVIDSPPH---QCVKIT-----VVAL 3196  
 QY 2871 SVT-YACEGYQLSLPAVFTCEGNS---WTGELPQCFPFECGD-----PGVPSR 2916  
 DB 3197 QLTSDDCQNY---LQLODSQGHNSFPQCGRNASAVPFFYSSMTAMVIFSGVNR 3253  
 QY 2917 GRREDRGSYR-SSVSFSCHPPLVVGSPRRFCQSDGTWSTGTPSCIDPTLTTCADGPV 2975  
 DB 3254 NSR--MSEPTYOIAOCNDRYHKAFCNLRSF---GMPDNYDNDXDCVTTLA---P 3299  
 QY 2976 Q-----FGIIONSGQVQGVSTVLFRCQXYL-----LQSTTTCILPN- 3013  
 DB 3300 QNHITSLPFSHGLIENSVE-----CRNDFLEVRNGSNSNPLGKYCGTLLPNP 3348  
 QY 3014 -----LTWSGTPDCVPHHCHOPETPTANVAGALDLPMSGY 3049  
 DB 3349 VFSQNNELYLRFKSDSVTSBRGEYIITWSSPSC-----GGTLYGDGCSFTSP--GY 3398  
 QY 3050 -----TLITPAR-----ASPS----- 3061  
 DB 3399 PGTYPNNTYCEWVLVAPAGLVLTINFYFISIDDPGDCVQNYLTLYDGPNNASSPSSGPYCG 3458  
 QY 3062 ---RVAPSTAPAR---RMAAGQARPSAMR 3085  
 DB 3459 GDTSIAPFVASNSQVFIKFIHADYARPSAFR 3489  
 RESULT 10  
 ID 060494 PRELIMINARY; PRT; 3623 AA.  
 AC 060494; 096RUG;  
 DT 01-AUG-1998 (Tremblrel. 07, Created)  
 DT 01-JAN-1999 (Tremblrel. 09, Last sequence update)  
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
 DE Intrinsic factor-Bi2 receptor precursor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OC NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97476251; PubMed=9334227;  
 RA Birt H., Verroust P.J., Nexo E., Hager H., Jacobsen C.,  
 RA Christensen E.J., Moestrup S.K.;  
 RT "Characterization of an epithelial approximately 460-kDa protein that  
 RT facilitates endocytosis of intrinsic factor-vitamin B12 and binds  
 RT receptor-associated protein."



```

Db      1693  RGRYCGTMBPBITFSALTLRFVSDSSISAGFHTTWTASVSACGCTFMAGIENSP 1752
Qy      1393  NYPQVYTSQOICLY-FVTVPKDYVVFQGPAPFHTL-----NDVEVHDGSHQSRLL 1444
Db      1753  GYDPIPPVNEBCVMNIISSPGNRL---QISFISFQLEDSODCSRPVEIRBGA---T 1804
Qy      1445  SLSGSHTEBESLPLATSNQV---LIKFSAGKLAPARGHFVYQAV-----1486
Db      1805  GHLVGRYCGNSPFLNYSIVGHITLMVRFISDSSGSGTFOATPKIKIGNINVTGKVA 1864
Qy      1487  -----PRTSAYO-----CSSVPEPRY-----1502
Db      1865  SPFMPENYHNNSYQWTVNVNASHVHGRILEMDIEIIONCYDKLRIDGPSIHARLIG 1924
Qy      1503  ---GKRLGSDPSVGAIVAFECNAGVALQGSPEICGLPVPGALAQW-NVSAPTCVP--- 1554
Db      1925  AYCGTQRESFSSSTGNSLTFHFYSDSISGK-----GFLBFWAVDAPDGLPTIAP 1975
Qy      1555  ---CGGNTERRGT-----ILSPGFPEPYLNSINCYMKIIVPEGAGIQIQAVSFVTEON--- 1605
Db      1976  GACGGFL--RTGDAPVFLFSPGMPDSYSNRVDCTWLLQAPDST-VELNIIISLDIESHRC 2032
Qy      1606  -WDSLEVPDGDADNTVMTLGSFGTVPALANSTNOUYLHFPSDISVSAAGFHELYKTVG 1664
Db      2033  AYDVLVRDGDNNIAQOLAVLCGREIPQPIRSTGEYMFIRFTSDSVTRAGFNASFH--- 2089
Qy      1665  LSSCPREPAVPSNGVKTGERYLNDVVSFOC-----EPGYALQGH-----AHISCM 1709
Db      2090  -KSGGGLVHADRGITITPKYPERYPSNLNCSMHVLSGLTIAVHFQPIQINGDSSCN 2148
Qy      1710  PG---TYRRMY--YPPPL-----CIAO-----1726
Db      2149  QGDYLVLRNGBDITSPPLGPPGNGHFCGSHASTLETSDNQMFVQFISDHSNEGQCFKI 2208
Qy      1727  -----CGGTV-----EEMEGVILSPGFQNYPSNMDCSMKIALPGBFAHIQF--LN 1771
Db      2209  KYEAKSLACGQNYIHDADAGVYTSNHNPHNYPBHADCTWIIAAPPETRIQIQFDRPD 2268
Qy      1772  FSTEPPN--HDYIEIRNGPYETSMMGRFSGSELPSLSLSTSHETTVFHSDHSONPGPK 1829
Db      2269  IEVTPNCTSNVLELRDGVDDAPILSKFCGTSLSQSGWMSGEMVYLRFSDNSPTHVGKR 2328
Qy      1830  LEQVAYVLEQCPREPPANIGVAGY-----NVGQSVTEPCLPQYOLT-----1873
Db      2329  AK---SIAQCGGRVPGQSGVSESIGHPLPYRDNLFCEWHLQGLSGHYLITISFEDFNQ 2385
Qy      1874  ---GHVPLTCQHGNTNRMD-----HPLP-----1893
Db      2386  NSSG-----CEKDFVEIMDNHTSGNLTGRYCGNTIPDSIDTSNTAVVRFVTDGSVTASG 2440
Qy      1894  -----KCEVPCCGNTSSNGVYVSPQSPYSSODCVWLITVPIGHVRL--NLSLQ 1945
Db      2441  FLRPFSSMEBCGGDLOGSIGTFTSPNYPNPHGRICEWRITAPBGRITLTMENLRRA 2500
Qy      1946  TEBDS--GDFTIWDGPOQTAPRLGVTRSM-AKTYQSSNOVLKFKHDAAT--GGIFA 2000
Db      2501  THSSCNNEHYIYANGIRSNSPLEKLCSSVNSNETKSSGNMTKVIFFTDGSRPYG--FT 2559
Qy      2001  IASAVAPLTKCPPTLLPNAEVVTENEBFNIGDIVARCLPGFTLVGN-----2048
Db      2560  ASYTSSEDAVCGGS--LFPN---TPRGNT-----SPGYDVRNYSRLNCEWILS 2604
Qy      2049  -----EILTCKLGYLOEGR-----PPICE 2069
Db      2605  NPNQGNSSISIHEDFYLESODCCPDVLEFRVG--DADGPLMRWLCGSPKPTLPIVIP 2661
Qy      2070  -----VHCPTNELL-----TD-----SNGVILSOSYGSYPOQOTCGWYRV 2106
Db      2662  YSQWVHFTNRVHEIGHAKYSFTDCGCIQIGSGVITSPNYPNADSLTHCWMLEA 2721
Qy      2107  EBDYNISLTVEXYFLSEKQ---YDEFELPDGSGQSPLLKALSGVNSAPLIYTSSSNGV 2162

```

---

```

Db      2722  PQGHTITLTSPDIEDIPHTTCAMDSTVYRNGSPSPSPITIGQYCGN-SNPRITQSGSNQLV 2780
Qy      2163  LRNSDHAANYRKGFKIRYSAPYCSLPRARLHGFILOQTSTOPOGSIHFCNAGYRLVGS 2222
Db      2781  VTFNSDHSIQGGGF-----YATNWTOT-----LCCGG---IFHS 2811
Qy      2223  MAICTHHPOGHILMSAIIPLCOALS--CGLEPAPKA-GMVFGLEYVGTAVYSCSEGYHL 2280
Db      2812  DNGTISPH---WPQNFENSRCSWTALTHKSKHLEISFDNNFLIPSGDGOCOSFVAY 2867
Qy      2281  OGAEBEATBEC-LDTGLMSNRNVPQCVPTCPDVSISVEHGRWRLIFETQYQOAOQLML 2339
Db      2868  WAGTEEVDRKALLATGC--GNVAPG--PYITSNFTAV-----FOSQEA- 2907
Qy      2340  ICDPGYTYTGQVIRCOANGKWSLG-----DSTPTCRIISGELPIPPNGHRIG 2388
Db      2908  ---PQGFSAFVSRCGSNFTGSPGVIISPNYKQYDNNMNTYV-----LEAN----- 2953
Qy      2389  TLSYVGATAI-PCSNQSVTLVGSRRVECMANGLMSGYRCLAGH-----CG--T 2435
Db      2954  PLSTVLLTFSHMLARSAVTGS---LVNDS---VHIIIRGYSVNSTPFAVYCGDEM 3003
Qy      2436  PEPPI-VNGHINGENYSYRGSVVYQCNAGFRLLGMSVRICQODHMSGKTPFCVPITCGHP 2494
Db      3004  PAPLTIAGVLLNFYSNBOITDF--GFK--FSYRI-----ISCGGV 3040
Qy      2495  GNPVNGLOQGNFNLNDVVKFVCNPGY-MAGSARSOCLASQWSDMLPTCRIINCTDPG 2553
Db      3041  FPFSSGI-----ITSPAYSAADYPMDMHCLYITVTSB-----3072
Qy      2554  HOENSVROVHASGPHRFSGTIVSYRCHNGFYLLGTPVLSCQGDGTWDRPRPQCLVSCG 2613
Db      3073  ---DKVIELKFS---DFDVVPBTSCHDVLAI-----YDGANTSD---PLIKRFGC 3114
Qy      2614  HEGSPPHSQMSG-----DSTYVAVVRYSCIGKRTLVGNSTRMGDLGHWGTSJPH 2664
Db      3115  SK-REPNVKSNNMSMLVFKTDSFOJAKGWSKSF--RQTL--GPOGGG--GYLTG--- 3163
Qy      2665  CSGTGVCVCGDPGPIPAHGIRLGDSPDPG-----TYWRFSCEAGHYLR- 2706
Db      3164  ---SNMTPASPDSDSNM-----YDKNLNCWIIIAIPVUNKVILHTFNTFALBAASTROR 3214
Qy      2707  -----GSSERTCOANGSMWSQPECCGVISCGNPGTSPSNARVVSDDLVSFSSSIY 2756
Db      3215  CLYDVYKLYDDSDSNANILA-GTFCGSTVPAPFISGNPLT-----VQISLDTL 3262
Qy      2757  ECRGEGYAT-GLLRHC--SVNGTWG---SDPECLVINCGDPGPIPANGRLGNDFRYNK 2810
Db      3263  E-REGFNATYTIMMPCGGTYNATWTPONISSP-----NSSDPVPVPS- 3304
Qy      2811  TVTYQVPGVYMMESHRSVLSCTDRITMNGTKPVCKALMCKPRPLINGKVVVGSDFPMWGS 2870
Db      3305  -----ICTWVIDSPPH--OOVKIT-----WVAL 3325
Qy      2871  SVT-YACLEGYQLSLPAVFTCEGNGS---WTGELPOCFPVFCSD-----PGVPSR 2916
Db      3326  QLTQDCTQYV---LOQDSPOGHGNSRFOFCGNANAVVVFVSSMSTANAVIFSGSVNR 3382
Qy      2917  GRREDRGESYR-SSVSISCHPPLVVLVGSPPRFQCSDSCTWGTGTOBSCIDPLTTCADGPV 2975
Db      3383  NSR--MSFTYQIADCNNDYHKAFGNLSR-----GMPDNYDNDDXCTVITLTA---P 3428
Qy      2976  Q-----FGIIONNSQGVQVGSVTLVFRQCKGYL-----LOGSTTRTCAPN- 3013
Db      3429  QNHTISLFFHSLIGIENSVE-----CRNDFLEVNRGSSNSNBLKRYCGTILFNP 3477
Qy      3014  -----LWSTGTPPDQVPHHCQBPPTPTHANVAGALDLPMSGY 3049
Db      3478  VFSQNNELYLRFKSDSVTSDRGEYIITWTSBPGC-----GGLYLDGDBGSFTSP--GY 3527
Qy      3050  -----TLTPPARB-----ASFS-----3061
Db      3528  PGTYPNNTYCEWVLVAPAGRLVITINFFYFISIDDPGQCVQNYITLTYDGPANASSPSGPYCG 3587

```

Qy 3062 ---RVAPSTADAP-----RMAAGQASRPASMR 3085  
 Db 3588 GDTSIAPFVASSNOVIFKFMADYARPSAFR 3618

## RESULT 11

Q9TUS3 PRELIMINARY; PRT: 3620 AA.  
 AC Q9TUS3;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Cud11n.  
 GN Name=CUBN;  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_Taxid=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=20021710; PubMed=10552972;  
 RA Xu D., Kozyraki R., Newman T.C., Fyfe J.C.;  
 RT "Genetic evidence of an accessory activity required specifically for  
 RT Cud11n brush-border expression and intrinsic factor-cobalamin  
 RT absorption.";  
 RL Blood 94:3604-3606(1999).  
 DR EMBL; AF137068; AAF14258.1; -.  
 DR HSSP; P35555; 1EMN.  
 DR GO:GO:0005509; F:calcium ion binding; IEA.  
 DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_CA.  
 DR InterPro; IPR006209; EGF\_like.  
 DR Pfam; PF00431; CUB; 27.  
 DR Pfam; PF00008; EGF; 4.  
 DR Pfam; PF07645; EGF\_CA; 3.  
 DR SMART; SM00042; CUB; 27.  
 DR SMART; SM00179; EGF\_CA; 4.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE; PS01180; CUB; 27.  
 DR PROSITE; PS00022; EGF\_1; 4.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS00026; EGF\_3; 5.  
 DR PROSITE; PS01187; EGF\_CA; 4.  
 DR EGF-like domain.  
 SQ SEQUENCE 3620 AA; 397432 MW; 00B041EB6AD07348 CRC64;

Query Match 8.6%; Score 1457.5; DB 2; Length 3620;  
 Best Local Similarity 21.4%; Pred. No. 6.5e-74;  
 Matches 679; Conservative 344; Mismatches 1070; Indels 1073; Gaps 117;

Qy 113 TCMKYSIDWFAA-----WSDHRYPCARWCDAMHLEGPSSIIISPNFILOYDNNACWITIT 167  
 Db 443 TCVGAVNAFSCCECTRFMTGFLCQIPQVCGSLSDMSGSFYSMDVGVYADVNCFWIR 502  
 Qy 168 ALNSKVKILAEFEEDLER---GYDTLTVGDGGDQDKTVLYWSQACSDSPHTPSR 223  
 Db 503 T-EDRKVLRIFFTFQLESVANNCHPEFLOI---HDGSSAALQIGRCGSLVLRH----- 552  
 Qy 224 IPESMSGDIMRQKKTVLEICRDISSDARSVKSPKPTSNVAVELVAPGTEIEGSGCDP 283  
 Db 553 --ELLS-----SNNAL----- 561  
 Qy 284 GIPAYGRREGSRPHNGDTLKECPAPFELVGKALITCQKNNWSAKKPCVFSCEFNFTS 343  
 Db 562 -----YHLVSEHRSRGRTI-----RMETQDECGGILMGTY-- 595  
 Qy 344 PSGVLLSPNYPEDYGNHLHCWMLILARPESRIHLAFNDIVER---QFDFLVIKDGATA 399  
 Db 596 --GSIKSPGYGVNRPGRDVCWQVVTSPDLITFTFGTSLSEHHDDCKDYLEIRDPFLY 653

Qy 400 EAPVLGFSGNQPLPSITSSGVARLEFQDHSNGKRPNTFTT----- 444  
 Db 654 QDPSELGFCCTTSLVPLQTTGPPFARVHFHSNQINDGPHITLYLSPSDLHCGANYDPE 713  
 Qy 445 -----FRHN-ECPPDGVNPGKRFQSDLOLG-----SSISFLDEG 479  
 Db 714 GLSSDLSGPFTHNRQC-----IYIKQPLBQIQVNTVHVELEGSSCSQSHIEVRDKI 769  
 Qy 480 FLG-TQSEITTCYKKEGSVW-----NSAVLRC-----EAPCGHLTSPGTTLSPPW 527  
 Db 770 LKGVACNETHLPHIKSIRNHIMWLKIDASILVASFPAYVQVACGSELTG-EGVIRSPFY 828  
 Qy 528 PGFYKDALSCAMVLEAPGVYIKTFPRFKTE-----VNYDTLEVRDRTISAPLIGVYHG 583  
 Db 829 PNVPYGERICKWTIHQPOSOVYINLTFAPGIESSAHCDTDYIEIGSSSILGSPENKRYCG 888  
 Qy 584 TQVQPLISTSNLYLLFSTDKSHSDIGFQRLRYETITLQSDHCLDPGIPVNGQRHNDIFY 643  
 Db 889 TDIPLFITSVNPLVYLVFKSSSTENGFMAKFS----- 922  
 Qy 644 VQALVTSCDSGYTLSDGEPLCEPWFQWSBALPSCALCGGFIQSSGTTLSFGPPDY 703  
 Db 923 -----AADLACGEILTESTGIIQSPGHNIX 948  
 Qy 704 PNNLCWTIETSHGKGVFTFHTFHE-----SGHDVLLITENGSFTQPLRLQTLGRPLA 759  
 Db 949 PHGINCTWHLIVQGHILHLIFRFHLEFHYNCTNDILEYDTSNTY-LGRYCKSIPR 1007  
 Qy 760 PISALGNFPAQVRFISDFSMSYEGNITFSEYDLPCBEPVAVYSIRGLQFGVQDT 819  
 Db 1008 SLTS---STNSKLIFPADSDLAIEGLINVEALDAS----- 1041  
 Qy 820 LTFSCFPGYRLGEGTARITCLGRRRLWSSPLPRCVACGNSVTGTGTLSPNFPVYNN 879  
 Db 1042 -----SACMEDYTGNSGTFTSPNPNVYPN 1066  
 Qy 880 NHECIYSIQTPGKGIQAKARPELSF-----GDVLKYDDGNMNAKLLGVSHSEBM 932  
 Db 1067 NMKCIYRIVETSQDILHFTNFLEALIGQCVADVEILRDGYETSPRLGYCGS-IP 1125  
 Qy 933 GVTLNTSSSLIMDFITDAENTSKEFLHFSFELIKCEDPGTPKFGYKVDHGFAGSS 992  
 Db 1126 PRLIHSNKLMLQFTSD-----FLGSG 1148  
 Qy 993 VSFSCDPPGYSILGSEBELLCISGERRVDRPLPTCVACGGTVRGVSGQVLSPGYPAYE 1052  
 Db 1149 -----PGFS-----AYWDGSL---TCCGGNIT-TPTGVFTSPSYPMPLY 1183  
 Qy 1053 HNLNCIWTIEAAGCTTGLHPLVDTBE---VHDVLRINDGPVRESGVLLKELSGPALPK 1108  
 Db 1184 HSSECYWLLKXSHSPLELEBDBHLEHNPCTLDYLAUVDDGSTSHLLSOLCGNKPX 1243  
 Qy 1109 DLHSTFNSVYLQSFDTDFTSKQGFALQF-----SVSRATSCNDPGIQRQNGSR 1156  
 Db 1244 VIRSTGSMFLKFTTDDQCGGFLAKYQOTCRNVVIVNKNYGLISEIHINPNPSDQRC 1303  
 Qy 1157 GDSWEA-GDSTVFOCDBGVALOGSABISCVKIE-----NRFWQPSPT----- 1199  
 Db 1304 NMTIQATGNTVNTVFLAFELHNHINCSYDLELYDPRRMGRYCGADMPTGTTGSKL 1363  
 Qy 1200 -----CIAFCGDLTGPSGVLLISPNYPRPYPRPKGCDKMAVTVYSPD 1239  
 Db 1364 QVLFYTDGVGHQEKQFQWQFHHCCGGBELGTGTSFSPGVPNTYPPNKKCIWITTAAG 1423  
 Qy 1240 VYIALVFNIFNLB---PGYDFLIHYDGRSLSPUG---SPYSGQLPGRIEISSNSLFL 1292  
 Db 1424 SSIGLTLHDVDEYHNAACNFDVLEVYGGPDPHSPRITIQLSQSSSNPMQVSSIGNELAI 1483  
 Qy 1293 AFRSDASVSNAGFVIDYTENPRESCFDPGSIKNGTRVGSCLKGSSVTYYCHGGEYVEGT 1352  
 Db 1484 RFKTDSSINGRGFVASHQAVP-----GG----- 1506

```

QY 1353 STLSCLIPDGPKNVNNRPVCTABCGGVYVSDGVLSPNVPOVYTSQGLCYFYVYVVK 1412
Db 1507 -----CGGIFQAPNKEIHSNVPYPSRYKRENTCSCWIVYER 1541
QY 1413 DYVEGQAFRITLAND-VVEYHDGSHSLSLSSGHTGESLPLATSNQVLIKESAK 1471
Db 1512 NHRILANTDPELQDSCITAYDGLSSTTRLASVCCRQOLTNPITSSGNSLFLRQSG 1601
QY 1472 GLAPRGFHYQAVPRTSATQCSVPEPRYGGKLGSPFSGALVPRPCNSGYALQSSPE 1531
Db 1602 PSRQGRFR----- 1610
QY 1532 IECLVPFGALQMVNSAPTCVPCGN-LTERRGITLSPGPPEBYLSLNCWMI-VVPE 1589
Db 1611 -----AQFN-----QVCGHILNTSPDTISPLPRAXPNQNSWYIQQRP 1653
QY 1590 GAGIQIOVVSFVTQDN-----WDSLSEVPDADNTYMLGSRGTTVPALLNSTNQYLHF 1645
Db 1654 FNHTTSLFDHGLSSTTCTQDPLEILDGDYDAPLGRYCGHSMPPHITSPSGALTLRF 1713
QY 1646 YSDISVSNAGFHELYKTVGLSSCEPAVPSNGVTKGRYLVDVVSFOCEPVALQGHAF 1705
Db 1714 VSDSRVNSDGFHATY-----AASSA----- 1734
QY 1706 ISCPHGTVRKNNYPPPLCIAQCGGTVEEMEGVILSPGPNGYPSNMDCKWIALPVGFGA 1765
Db 1735 -----CGGTFHMAEGIFNSPGYPEYPSNVECVNINYSPPGNRL 1773
QY 1766 HIOFLNFTSP-----NHDXIEIRNGPYETSRMRKRESGSLPSLS-TSHETTVYFHS 1820
Db 1774 QLSITITQLEDSQDCSDRPFVEKREG-NATGHLVGRYGVNLPLNYSIVGHILMIRVSD 1832
QY 1821 HSQNRPGFK----- 1830
Db 1833 GSGGTGFQATFKTIFGNDNIVGTHKIASPMPGRYHNSNQWIVNVAQTQVHGRIL 1892
QY 1831 EYQAYEIOEC-PDEPERANGI--VKGAGYVNG-----QSVTECLPEYQULTGHP 1876
Db 1893 EIDIEGASQCYKRLRYDGLGHSRLIGYCGTQTTSPSSRNNSLTFQPSDSITGKG 1952
QY 1877 VLTQHGNTNRMDHPLEKCEV-PCGGNITSSNGTV--YSPGFSPSPSSQDDCWLTVPI 1933
Db 1953 FLEWEPVANS-GGRLPTITAGACGFLRTGDAVFLFSPGWPEBSYNSADCWLQAP- 2010
QY 1934 GHGVRNLNLSLQTEPSG---DETTIWDGPOQAPRLGVFTRSMARKTVOSSNQYLAF 1989
Db 2011 DSTVELNITSLDIEAGRTCDYDKLVIIRDGDSNLAPQAVLCGEIRGPIRSTGEVFIHF 2070
QY 1990 HRD-AATGTFALAF-----SAVPLTKCP--PPTILPNAE-----VVT 2024
Db 2071 TSPFSTITGAFNASFHKSCGYLHADRGITTSPOYEITYSPMLNCSMHVAVOSGLTIAVH 2130
QY 2025 ENEEFNI--GDIYRVRCLEPFTLV--GNEILLTCKLGTQLQ--FEGPPPIEYHOPFNE 2076
Db 2131 FEQPFQIPSD---SSCSQGDYLVLNKNGPIYSPPLGPGKRNHFGSGSPSSTLFTSDQ 2187
QY 2077 LL-----TDSGTVILLQSGYPSYPOFQTCMWL 2103
Db 2188 MFVQFISDSNGSGQFKIYEAKSLACGNIVYHDVNSAGYVTSPPHPNNYVQHDQMWL 2247
QY 2104 VRVEPPIYNSLVEYELSEKQYDE-----EIFQPSQSLALALSGNTSA 2150
Db 2248 IAPRPGKLIRVOPE-----DQFNIEETPNCVSNYELRGGVDSNAPLAKLQGR-SL 2298
QY 2151 PLIVTSSNSVYLKRWSSDAHYNRKGFIRYSAPYCSLPAPLHGFILGQTSTQPGSILHF 2210
Db 2299 PSSQSSSGVWYLRFPSSDSSNQVGFKIKYATAQC----- 2333
QY 2211 GCNAGRLVGHSMALITRHPQGHMWSBALPQLALSCGLPAPAKNGVFGKEYTVGTGA 2270
Db 2334 -----GGRVYVGSGI-----ESSGYPTLPYRDNSTFCMHLKXPSG 2369
QY 2271 VYSC--SEGVIHQAGAEATATACLDGTGLMSNR--NVPPQCVVTCPD----- 2312

```

```

Db 2370 HYTLTFEDHLNONGSCCKDFVE--IWENHTSGNLGGYCCNTTIPDSIDTSSNALVRF 2427
QY 2313 VSSISVEHGRWRLIFETQY-----QFOQLMLICPGYTYTGQVIRCOANGKMSLGST 2367
Db 2428 VTDSGYTAGSGLRFRFESSMEACGGELQCPGTGFTSPNY-----PNRN 2469
QY 2368 PTCRIISCEBLPPPNHIG-TLSYVGTALFSCNSGYTLVGSRYRECMANGLSGSEV 2426
Db 2470 PHGRV--CEWRIMVOGRRITLTFFNNLRLEAHPSCVSEHVTI-----FNGIRNNSDQ 2519
QY 2427 RCLAGCGTPEPIVNGHNGENYSYRGSVYVQCNAGFRILGNSVRIQQDDHMSGTPPC 2486
Db 2520 --LEKICGS-----VNASSIKSGNTMKVFTDSDRPGGSAT-----YTSSEDAVC 2567
QY 2487 VPTTCGHPNPNVNGLTGQNFNLNDVVKVPCNPGYMAEGAASQCLASQMSDMLPTCRI 2546
Db 2568 -----CGSLTHPPEGN-----FTSPGY--NGVSNYS-----RN 2593
QY 2547 INC-----TDPGQENSVRQVHASGPHRFSFGTTVYRCNHGTYLLGTPVLSQGD- 2597
Db 2594 LNCWETLSNP--NOGNSIYIHF-----EDFYLSHQ--DQDFVLBR 2633
QY 2598 -GTWDRPREPQCLVSCGHPSPPHSGMSGDSYTVGAVVRYSCIGKRTLVGNSTRMCGIDG 2656
Db 2634 VGNADGRLMKRL--CG-PSKP-----IVPLVIFPEVWTHFVTHNEHVEHYGF- 2677
QY 2657 HMTGSLPHCGSTV--GYCCDGPPIPAHGRIGLQSDFDGYVRFSCEA--GHVLRGS-- 2708
Db 2678 HAEVFTPCGGIQLDSGVIASPNYPA-----SYDLTLCMSWLEAPQGFITLTFTSD 2730
QY 2709 -----SERTQANGSSGQPECGVISCNPGTP-----SNARVVSQG--LVSS 2752
Db 2731 FIDIEHATC--AMDSVVRNG-----GSPSPITIGYCGTSNPFTQSGSNQVLVIFNS 2782
QY 2753 SIVECREGYAT-----GLLSRHCYNGT-----WTGSDPECLVINGCPGPIAN 2798
Db 2783 DHSVQ--NGGFATVMTQTLGCGGILH--SNGNIRSPHMQNPPE--NSCSMTVITHE 2836
QY 2799 GLRLNDNRNKTVTV--QVPGYMMESHRVSLCTKDTNTNGTKPVKCALM--CKP 2852
Db 2837 SKQLEISFDNNFRIPSGDQCONSIV-----KVMAGTEVVASLSTATGC-- 2880
QY 2853 PPLIPNGKLVGSDPFMWSSTVYACLEGYQ-----LSLPVYTCGCGN 2893
Db 2881 -----GNVARGSLITRNVITAVFQSOETPAQGFSAFVSRCGN 2920

RESULT 12
ID 0723G3
AC 0723G3 PRELIMINARY; PRT; 1441 AA.
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Hypothetical protein DKFZp686i10142 (Fragment).
GN Name=DKFZp686i10142;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Human colon endothel primary cell culture;
RA Koehler K., Beyer A., Mewes H.W., Weil B., Amdt C., Osanger A.,
RA Podo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: Contains 3 EGF-like domains.
DR EMBL; BX537918; CAD97901.1; -.
DR InterPro; IPR000742; EGF_2;
DR InterPro; IPR006209; EGF_like;
DR InterPro; IPR006210; IEGF;
DR InterPro; IPR000436; Sushi_SCR_CCP.

```

DR Pfam; PF00008; EGF; 3.  
 DR Pfam; PF00084; Suelb; 22.  
 DR SMART; SM00032; CCF; 22.  
 DR SMART; SM00181; EGF; 3.  
 DR PROSITE; PS00022; EGF\_1; 3.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS00026; EGF\_3; 2.  
 DR PROSITE; PS00923; SUSHI; 22.  
 KM EGF-like domain; Hypothetical protein.  
 FT NON-TER 1  
 SQ SEQUENCE 1441 AA; 157497 MW; BA6B04EC59962A22 CRC64;  
 Query March 8.3%; Score 1416; DB 2; Length 1441;  
 Best Local Similarity 25.1%; Pred. No. 4.4e-72;  
 Matches 407; Conservative 179; Mismatches 591; Indels 444; Gaps 52;  
 QY 1668 CPEPAVNSGVKTERLYLVNDVVSFOCEPGYALQGHAIHSCMPCTVRMNPPLCTAOC 1727  
 DB 16 CGEPSPIMNGYASGNSYSGAMVAVYSCNKFYIEGKKSTC--BATQMSPIPTC---- 69  
 QY 1728 GGVTEEMEGVLSGPGFQNTSPSNDCKKIALPVGFQAHIOFLNFTSEPHNDYIEIRNGP 1787  
 DB 70 -----HPVSCG-----EPP-----KVENGF 84  
 QY 1788 YETSRMMGRFSGSELPSLSTSHETTYPHSD--HSQNRPGFK--LEYQAVELQ----- 1838  
 DB 85 LE-----HTTGRIESEVRVYQCNPGYKSGSPVFCQANRMHS 123  
 QY 1839 -----BCPDDEPPANGIVRGAGYVQSVTFECLPGYOLTGHPVLTCOHGTNRMD 1889  
 DB 124 ESLPMCVPLDCGKRPPIQNGFMKGFNEFVSGKVOFCNEGVELVDGDSMTQ--KSKMN 181  
 QY 1890 HPL-PKC-----EVPCGGN-----ITSSNGTVYSRQFSPYSSQDCWLTIVPIGKG 1936  
 DB 182 KKSNNPKMPKACRPPPLENLVLKELTTEGVV-----TFSCKE-----GH- 223  
 QY 1937 VRLNLSTLQTEPSQDFITIMDGPQGTAPRLGVFTSRMAKKTIVGSSNQVLKPHRDAATG 1996  
 DB 224 VLQGPVVKCLPSQO-----NW----- 240  
 QY 1997 GIFAIASAVPLTK---CPPTILPNAEVTENEENIGDIVRYCLPGFTLVGNEILTC 2053  
 DB 241 -----DSPFVCKIVCTPPPLI-SFGVPISSALHFGSVYKXSCVGFPLRNSITLIC 292  
 QY 2054 KL-GTYIQFEGRRPIC-EVHCPTNELTDSGTGVLSSQSPESYQFOTGCMWLVAVEDYN 2111  
 DB 293 QPQGTW---SSPLPECVPECPQPEEIPN--GIIDVOGLAYSTALYTIC-----KPGFE 341  
 QY 2112 ISLVEYFLSEKQYDFEIPFGPSGQSPILKALSGNTSAPLIIVSSSNVYLKMSRHAY 2171  
 DB 342 L-----VGNITTLGCGENGHW----- 356  
 QY 2172 NRKGFKIRYSAPYCSLPAPLHGFILGQTSYTOPGSIHFGCNAGYRLVGHSMALCTBPQ 2231  
 DB 357 --LGGKPTCKAIECLKREKLING--KFSYTDLHYGOIYITSGNRRFLREGPALLCLETGD 413  
 QY 2232 GYHLMSDALPLQALSGLEAPAPKXGMYFGEKYTVGTAKVYSGEGYHLDAQADATACL 2291  
 DB 414 ---WDVAPSCNAIHCDSPQIENGFEQADYSGAIIITSCPPGF--QVAGHAMQTCSE 467  
 QY 2292 DTGIMSRNVPQOVPTC-----PDVSSISVEHG----- 2321  
 DB 468 ESG-WSSS--LPTCMPIDCGLRPHIDRGDCYKLDKDDQYFEQEDDMEEVRYVTPHRYHL 524  
 QY 2322 -----RRRLIFE-----TOYFOAQLMLICDPGYUYTGQVIRICQANGKSLDSTPTC 2370  
 DB 525 GAVAKTMENTKESATHTSSNLYGTWVSYTCNPGYELLGNVLLCQEDGTMN--GSAFSC 582  
 QY 2371 RIISCGELIPIPNGHRIGTLSVYATPAIFSCNSGYTLVGSFVRECMANGMLSGSEVRCLA 2430  
 DB 583 ISICCDPLTAPENGFLEFTEKSMGSAVOYSCPKPHIILAGSDLRCLERKRSKASPRCEA 642  
 QY 2431 GHCTPPIVNGHNGENYVRSRVVQCNAGPRLIGMSVRLICQDDHMSKTPFCVPIIT 2490

DB 643 ISCKKPPVNWNGSISKSNTYTLSTLYECDPGVLTNGTERTKQDDKNMDEDEICLPVD 702  
 QY 2491 CGHPGPNVNGLTQGNQFNLDVVKFVPCNPGYMAAGARSOCCLASGQSDMLPTCRITNCT 2550  
 DB 703 CSSPPVANSQVAGDEYTFQKEIETHCNEGFLLEGARSRYCLANGSWSGATPPCVPRCA 762  
 QY 2551 DPHQENSVPQVASHRHSFGTYSYRCNHPFYLLGTVLSQCGGCTMDRRRPOCLV 2610  
 DB 763 TPQLANGVTE---GLDYGFMKEVTFHCBEGYLHGAFKLTQSDGNMDABIPLCXPV 817  
 QY 2611 SCGHPGSPHNSQSGSYTVGAVVRYSCIGKRTLVGNSTBMC----- 2652  
 DB 818 NCCEPBDLALGFNGFNGFHHGHQYQCFPGYTLHGNSRCLNSGSSSSPCLPCRC 877  
 QY 2653 -----GL-----DGHWTGSLPHCSGTSYVC 2673  
 DB 878 STPIVEYGVNGTDFDCKAARIQCFKGLILGISBITCBADQMSGSPHCEHTS---C 934  
 QY 2674 GD-PCIPAHGIRLGDSPDCTWRFSCEAGHYLRGSEERTQANGSSSQPBGVATSCG 2732  
 DB 935 GSLPMIPNADISETSWKE--NVITYSCRSQYVIQSSDLICTEKGVMSQPYPCCEPLSCG 993  
 QY 2733 NPGTSSNA----- 2740  
 DB 994 SPSPVANAAYATGEAHYTESVYKRLCEGYTMDITDTITTCQKDRMPERISCPKKCPL 1053  
 QY 2741 -----RVFSDGLVFSSTIVYECREGYATGLSRHCSVNGTWG--SDPECLVINCGD 2792  
 DB 1054 PENITHTLVHGDFFSVVRQVSVCABEGYFEGVNVISQLDGTWEPFSDSCSPVSCGK 1113  
 QY 2793 PGIPANGLRGNDPRNKYTYTQCVPRYMMESHRVSLSTKHRTNNGTPVYKALMCKP 2852  
 DB 1114 PESPEHGFVVGSKYTFESTIITQCEPDEYLEGNRERY--QENRQMSGVAALICETRCET 1171  
 QY 2853 PPLIPNGKVVGSDPFMMSSYTYVACLBEYQSLPAVFTCEGNSWTGELPOCFPVFCGDPG 2912  
 DB 1172 PLEFLNCKADIERKTGTPNNVYSCNKGYSLEGSSEAHCTENGTHSHVPLCKRNPCCVPF 1231  
 QY 2913 V-PSRGRREDRGFSYSSVSFSCHPPLVLVGSPPRRFQSDGTWSTGTOPSCIDPITLTTCAD 2971  
 DB 1232 VIFENALLSKEFVQDVQNSIKCREGFLQGHGIIITCNPDETWQTSKAC---EKISCGP 1288  
 QY 2972 PGVPQEGIQNNSQGYQVGSIVLPRCQGYLLQSTTRTCLPNIJWSTPPDQVPHHROP 3011  
 DB 1289 PAHVENAIARGVH-YOYGDMITYSQVYMLBEFLRSVCLENGWT--SPDIC-RAYCRFP 1345  
 QY 3032 ETPHANVGLDLP-----MGYTLTPARRASPSRVAFTAPARMAAGQASRPSA 3083  
 DB 1346 ---CONGGICQRPNACSCPEGMGRLC-----EPPICILPCLNGRCVAPYQCCPCPG 1395  
 QY 3084 W 3084  
 DB 1396 W 1396  
 RESULT 13  
 ID Q8NB79 PRELIMINARY; PRT; 1497 AA.  
 AC Q8NB79;  
 DT 01-OCT-2002 (Tremblere1.22, Created)  
 DT 01-OCT-2002 (Tremblere1.22, Last sequence update)  
 DT 01-MAR-2004 (Tremblere1.26, Last annotation update)  
 DE Hypothetical protein FLJ90754.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Iisagai T., Ota T., Nishikawa T., Hayaashi K., Otsuki T., Sugiyama T.,  
 Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,

Yamamoto Y., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,  
 Maehara Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,  
 Hatohori A., Okumura K., Iwayanagi T., Nishimura K.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 3 EGF-like domains.  
 DR EMBL: AK075235; BAC11489.1; -.  
 DR HSRP: P00740; IEDM.  
 DR GO: GO:0004872; F:receptor activity; IEA.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR006209; EGF-like.  
 DR InterPro: IPR006210; IEGF.  
 DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam: PF00009; EGF\_3.  
 DR Pfam: PF00084; Sushi; 23.  
 DR SMART: SM00032; CCP; 23.  
 DR SMART: SM00181; EGF; 3.  
 DR PROSITE: PS00022; EGF\_1; 3.  
 DR PROSITE: PS01186; EGF\_2; 2.  
 DR PROSITE: PS00026; EGF\_3; 2.  
 DR PROSITE: PS0923; SUSH1; 23.  
 DR EGF-like domain; Receptor.  
 KW EGF-like domain; Receptor.  
 SQ SEQUENCE 1497 AA; 163802 MW; 35B645633C3733B CRC64;

Query Match 8.3%; Score 1415; DB 2; Length 1497;  
 Best Local Similarity 24.9%; Pred. No. 5.2e-72;

Matches 405; Conservative 183; Mismatches 615; Indels 426; Gaps 49;

QY 1486 VPTSTATQSSVPPRYG---KRLGSDPSVGAIVFPCNSGYALQSGPEIECTLPVFGALA 1542  
 DB 1 MPRCIAHFCEKPEFSVSITLESVAKAFAGSVVSFCMEGFVLTNSAKIECH---RGC 56  
 QY 1543 QMNVSAFTCVPCCGNLTERGRTLLSPGPPEPILNSLNCVWKIVPEGAGIQIIVSVPT 1602  
 DB 57 QMN-----PSPM--SIQC----- 67  
 QY 1603 EONWDSLEVPDADNTVTMLGSEFGTVPALLNSTSNQVLYLHYSDISVAGFHELYKT 1662  
 DB 68 -----IPV----- 70  
 QY 1663 VGLSSCEPVPNSGVKTERVYLVNDVVSFOCEPFGVALQGHAIISCMPTVRKMNPPPL 1722  
 DB 71 -----RCCEPSPIMNGVAGSGNYSFGAMVAVSCNKGFIYIKGKKSTCEAS--GQWSPPIPT 124  
 QY 1723 C-LAOCG-----GTYEHEGVIL-----SPGEPGVPYNNMCCSKIALPVFGAH 1766  
 DB 125 CHVSGCEPPEKXENGFLHTTGRIFSEVARYQCNPGYK-----PVGSPVF 169  
 QY 1767 IOPLNFTSEPNHDYIEIRNGPYETSRMGKRFSGSELPSLSTSHETTVYFHSDHQNP 1826  
 DB 170 VCHAN-----RHMSESPLMWVPL----- 188  
 QY 1827 GFPLRYQAYELQCEPPEPPANGIVRGAGVNGQSTPECLRGYQLTGHVPLTQHGTR 1886  
 DB 189 -----DCGKRPPTQNGFMKGNEFEVSKQVFCNDEYELVAGDSWTCQ--KSG 234  
 QY 1887 NMDHPL-PRC-----EVPQGN-----ITSNGTVYSGPSPVSSQDCVMLITVPI 1933  
 DB 235 KMMKSNPKCMKAPKCEPPLLENQVLYKELTEVGV-----TFSCGE----- 277  
 QY 1934 GHGCVRLNLLQTESGDFPTTWDGPOQTAIRLGVTRSMAKTVQSSNQVILKFHDA 1993  
 DB 278 GH-VLQGPSTVLKCLPSEQ--WN----- 296  
 QY 1994 AAGGIAINAFSAVPLTK---CEPPTLRPAEAVTENEENIGDIYRIKLPFTLVGNEI 2050  
 DB 297 -----DSFPVCKIVLCTPPPLI-SFGVPIPSALHFGSTYKYSVGVGFFLRGNS 345  
 QY 2051 LITCKL-GTYLOEGEPPIIC-EVHCPTELLTSTGYLSQSYRGSYPOGOTCSMLVRYEP 2108  
 DB 346 TLQCPDGTW---SSPLPECVPECPQPEIEN--GIIDVRGLAVSTALYTC-----KP 394  
 QY 2109 DYNISLITVYFUSEKQYDEFETFDGSSGSPILKALSGNYSAPLLIVTSSNSVYLWMSD 2168

DB 395 GFEL-----VGNITTLGCGNGW----- 412  
 QY 2169 HAYNRGFKIRISAPYCSIPRAPLHGFLIGOTSTPGSGSIHFGCNAVRLVGHSMICTR 2228  
 DB 413 -----LGGPRTCALECLKPEKILNG-KFSYTDLHNGQIVTVYSCNNGFLRSGSALTJCE 466  
 QY 2229 HPQGYHLMSEAIPLCOALSGLPEAPKNGMVGEKTVCTKAVYSGSEGYHQAQGEAFA 2288  
 DB 467 TGD---WVDAPSCNAIHCDSPQPLENPFVAGADYSALITITSCPPG--QVAGHANO 520  
 QY 2289 ECLDTGLMNRNVPPOCVPTC-----PVSISVYEHG----- 2321  
 DB 521 TCEESG-WSSS--IPFCWPIDCGLPPIIDFGCTKLKDDQYFEGEDDMVEVYTPHP 577  
 QY 2322 -----RWRLIFE-----TOYFOQALMICDPGYTYTQRYIRCOANGKMSIGST 2367  
 DB 578 YHLGAVAKTWENTKSPATHSSNFIYGTWVSATCNGPGLLNPLVICOEDG--TRNGSA 635  
 QY 2368 PTCRIISCGELPIPPNGHRTGLSVYGATAIPSCNSGYTLVGSRYREGANGLWGSSEVR 2427  
 DB 636 PSCISIEDOLPTAPENGFLRFTETSMGSAVQYSCKPGHILVSGRLCLENKMKSGASBR 695  
 QY 2428 CLAGHGTPEPIVNGHINGENYSYRGSVVYOCNAGFRLIGMSVRIQQDHHWSGXTPECV 2487  
 DB 696 CEAISSCKENPVMNIGSNTYTLSTLYEEDPGYVLNGTERRFCODDKWDEDEPICI 755  
 QY 2488 PITCGHPGNVVLGQGNQFNLDVYKFCNPGYVAEBAASOCCLASQMSMLPTCKII 2547  
 DB 756 PVDRSSFPVSANGVQGDGYETFOKEIETYNCEGLFBEARRRVCILANSWSGATDCAV 815  
 QY 2548 NCTDGHQENSVYQVHAAGPHRFSGTIVSYRCNHFYILGTPLYVSCGDDGTDRPRQC 2607  
 DB 816 RCATPQLANGYTE-----GLDYGMKEVTYHCHGCVILHAPKLTQSDGNWDEIILC 870  
 QY 2608 LLYSGCHGSPHSGMSDSTVGAIVARYSCIGKTLVGNSTRMCGLDGHTGSLPHCSG 2667  
 DB 871 KPVNCGPEPDLAGHPNGFSFIHGCHIQQCFPGYKHLGNSRRCLSNWSGSGSPCLP 930  
 QY 2668 TSVGVCGPPIPAHGIRLGDSEFDPQTVWRFCSEAGHYLRGSSERTCOANGSMGSGQPCG 2727  
 DB 931 FR---CSTPVL-KYGVNGTDFDGGKARICCFKGLLGSLETTCADGQWSSGFPHCE 986  
 QY 2728 VISGNPETSNAHVPSDGLVFSSSIIV-YECREGYATGLLSRHCSVNGTWTGSDPECL 2786  
 DB 987 HTSCGS--LPMIPNAFISSETSMKENVITYSCRGYVYQSGSDILCTEKGWMSQPYRCE 1044  
 QY 2787 VINCGDPIRANGRLGNDPFRYNKTVYQCPYGMASHRSVUSCTKRDYTNKTPYCK 2846  
 DB 1045 PLSGSPSVANAVATGEAHTYSEVKURCLEGYTMDLT-TDPTCKDGRMPERISCS 1103  
 QY 2847 ALMKRPPLIPNGKVVQSDFMWGSVTVYACLEGYQLSLPAVFTCEGNGSMTGELP---Q 2902  
 DB 1104 PKKCPPLPENIHLVHDDPSVNRQVSVSCAEGYTFEGVAINISVQLDGTW--EPPFDES 1161  
 QY 2903 CFPVPCGDPGVPSRGREDGFSYRSVSSCHPPLVVSPPRRFCQSDGTWSTGPSCTI 2962  
 DB 1162 CSPVSCKPESEPHGFVGSKYTFESTIIVQCEGYELBEGRREVRVCOENQMSGVAIICK 1221  
 QY 2963 DPTLTTCADRCVPOFGIIONNSQGVQVGYTLVFCQKQYLLQSGTTRCLNLTMSGTPPD 3022  
 DB 1222 ETRCETOLEPLNGADIENT---TOPNVVISCNKYSLSLEGSEAHCTENGTSHPVPL 1277  
 QY 3023 CVPHHCRP 3031  
 DB 1278 CKPMPCPVP 1286

RESULT 14  
 070244 PRELIMINARY; PRT; 3623 AA.  
 070244;  
 01-AUG-1998 (TREMblrel. 07, Created)  
 01-JAN-1999 (TREMblrel. 09, Last sequence update)

DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
 DE Intrinsic factor-B12 receptor precursor.  
 GN Name=CUBILIN.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eulacerta; Rodentia; Sciurognathi; Muridae; Rattus.  
 NX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96148073; PubMed=9478979;  
 RA Moestrup S.K., Kozyraki R., Kristiansen M., Kayen J.H.,  
 RA Rasmussen H.H., Brault D., Pontillon F., Goda F.O., Christensen E.I.,  
 RA Hammond T.G., Verroust P.J.;  
 RT "The intrinsic factor-vitamin B12 receptor and target of teratogenic  
 RT antibodies is a megalin-binding peripheral membrane protein with  
 RT homology to developmental proteins."  
 RL J. Biol. Chem. 273:5235-5242(1998).  
 DR EMBL; AF022247; AAC71661.1; -.  
 DR PIR; T08618; T08618.  
 DR HSP; P00740; 1EDM.  
 DR GO; GO:0005559; F:calcium ion binding; IEA.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro; IPR000152; AaX\_Hydroxyl\_S.  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_Like.  
 DR Pfam; PF00431; CUB; 27.  
 DR Pfam; PF00008; EGF; 4.  
 DR Pfam; PF07645; EGF\_CA; 3.  
 DR SMART; SMO0179; EGF\_CA; 4.  
 DR SMART; SMO0179; EGF\_CA; 4.  
 DR PROSITE; PS00010; AaX\_Hydroxyl; 3.  
 DR PROSITE; PS01180; CUB; 27.  
 DR PROSITE; PS00022; EGF\_1; 4.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS00026; EGF\_3; 6.  
 DR PROSITE; PS01187; EGF\_CA; 4.  
 DR EGF-like domain; Receptor; Signal.  
 FT SIGNAL 1 20 Potential.  
 FT CHAIN 21 3623 Intrinsic factor-B12 receptor.  
 SQ SEQUENCE 3623 AA; 398984 MW; 39FB792AC6545240 CRC64;  
 Query Match 8.3%; Score 1404.5; DB 2; Length 3623;  
 Best Local Similarity 20.6%; Pred. No. 7.1e-71;  
 Matches 695; Conservative 378; Mismatches 1083; Indels 1219; Gaps 128;  
 QY 101 NEGYDLOGSKRITCMKVSMDPAWSDHRCVCRAMCDALHGRSGIIT--SPNFIQYDN 158  
 DB 443 NGRCICIGINGFTC----DCTSSWTGYCOTPOAACGILSGTGTAYHSPN--DTYIH 496  
 QY 159 NAHCWITITLNPASKVTKLAFESFDLE-----RGYDTLVGSGGQDQDQKTVLYMSQA 212  
 DB 497 NVNCFWIVRT-DEBKMVLHVTFTFDLSASNCPRY--LQIHGDSASDPIRGY----- 548  
 QY 213 CSDSPHPIPGSKIPESMGDIWKOKTVLEICRDISSDARSGSVKSKPKSNVAVELVAPG 272  
 DB 549 C-----GSRPFGI-----SSANALYFHLYS 570  
 QY 273 TEIEGSCGDPGIPAYGRREGSRFHGDTLKEGCPAFELVQKAITQKKNQMSAKKP 332  
 DB 571 EYISGR-----GFTA-----RREAKLPE 589  
 QY 333 CVFSCFNFSTSPSGVLSPNYPEDYGNHLCVWLILARPESRIHLAFNDIDVEP---QF 388  
 DB 590 CGGILTQNY-----GSITSFGYRGNVPRGDCVQVNVNPNLSLFTFTGTLBSHNDCK 645  
 QY 389 DELVIKGGATAEAVLGTFSGNQLPSSITSSGHVAREFQDHSITGKRGFTITTT--- 444  
 DB 646 DYLEIRGGPFHQDVLGKFTSLSTPLKTTGPAARHIFHSDSETSKGFIITLYLTQSD 705  
 QY 445 -----FRHN-----ECPDPGVVNGKFGDGLQSGSISF 474

DB 706 LDCCGNTDTDELLPPLSGPFSHRQCVYLITQAGQEQIVINFHVELESQWGCSTHY 765  
 QY 475 L-----CDEGLGQSGSTITTCVLEKGSVW-----NSAVLR--EAPCGHLTSPS 519  
 DB 766 IEVGDHSLLRKI CGNLTLPISVSNSKVRIRLRDALVQKASGRADYQVACGMLNG-E 824  
 QY 520 GTILSPGMPFYKDALSCAVVIEAOPGYIKITFDPRK-----TEVNDTLEVRDGRYSA 575  
 DB 825 GFRSPFYPMAYRERRCRRTTISOPKQVVLNFTDFOIGSSASCDTDYIEIGPSSVLGS 884  
 QY 576 PLIGVNGTQVPOFLISTSNVLLFSTDKSHDIFQLEVEYITTLQSDHCLDPGIPVNG 635  
 DB 885 PGNKFCSSNIPSPITSVNILLVTFVKSSSMENRGFTAFSSDKLE----- 931  
 QY 636 QRHNDPVGALVTFSCDSGTLSDGEPLECEPNFQMRALPSCALCGPFGSSGTL 695  
 DB 932 -----CGEVLASTGITE 944  
 QY 696 SPGPDPFNNLNCWTIETSHGKVPFTEHTHLE-----SGHDYLLITENGSTOPLRQ 751  
 DB 945 SPGHFNVPYPRGNCNMVAVVQRGOLIRLESSTFLEFHYVCTNDYLETITLAQTF-LGR 1003  
 QY 752 LTGSRPLPAPISAGLYGNFTAQVRFIIDPSNSYEGFNITFSEYDLEPCBEDEVAPYSIRKG 811  
 DB 1004 YCGKSIPPSLTS--NSNSIKLIFVSDSALAHGFSINYEALD-----ASSV--- 1047  
 QY 812 LQGVGDTLTFSCEPPYRLEGTAITLCLGRRRLMSSPLPRCAECCNSVTGQTLTSP 871  
 DB 1048 -----CLYDTDFNGMLSSP 1062  
 QY 872 NFPVNNNNHCTYSIOTQPGKGIOLKARAFELSE-----GDVLKYDDNNNSARLIGV 925  
 DB 1063 NFPNNPSSNNECTYRITVGLNQOLALHFTDFTLEDYFGSCQCVFELRDGEGYELPVG 1122  
 QY 926 FSHSEMMGVTLNSTSSSLMDFITDAENTSKGEFLHPSFELIKCEDPGTKRGYKVHDE 985  
 DB 1123 YCGS-VLPPTLIHSNKLWKFKSDALATKGSAYWD----- 1159  
 QY 986 GHPAGSSVSPSCPGVSLRGSBEILCLSGRRRTWDRPLPTVABCCGTATGEVSGOVLSP 1045  
 DB 1160 -----GSSGT-----CGGNLT-TPTGVLTSP 1179  
 QY 1046 GYPAPYEHNLNCWTIEAAGCTIGLHFLVPTDEYV-----DYLRTMDPVSQVLKE 1100  
 DB 1180 NYPPPTYSSECTWRLEASHGSPPELEFQDFLHEH-HPSCSLYLAVFDPTTNSRLIDK 1238  
 QY 1101 LSGPALPKDASTFNSVLOFSTDFTSKGFALIOF-----SVSTATSCNDP 1148  
 DB 1239 LCGDTTPAPIRSKNDVLLKLRTDAGQGGGFEINFRQCDNVIVNKTGILESINYPN 1298  
 QY 1149 IPONGSRSGDSWEA-GDSITVQCDPGYALOGSAEISCVKIE-----NRFWQSPPT 1199  
 DB 1299 PYDKNQRCNMTIQTATGNTVNTYFLGFDVESYNNCSFDYVELYDGPQMMGRYCGNNMP 1358  
 QY 1200 CMAP-----CGGLTGSSGYLSPNYPEPYPKQEC 1231  
 DB 1359 GATTSQHLVLFHTDINGSEKGFQKQWFTHGSGGSGTAGSFSSPGYNSYPHNECT 1418  
 QY 1232 WKYTVSPDYVIALVFNIFNLE-----PGYDLATIDYDGSLSPLIGSYG--SOLPRIE 1284  
 DB 1419 WNIIRVAPGSSIQLTIDFVBYHTSCYBDELEIYAGIDRSPRLAOLCQSPSPANPQVS 1478  
 QY 1285 SSSNSLFLAFRDSAYSNAGFVIDYENPRESCFPDPSIKNGTRVSGDLKLGSSTVYYCH 1344  
 DB 1479 STGNELAIRKXTDSTLNGRFGFNASWRAVP----- 1507  
 QY 1345 GGYEVBSTSLSCILGPDGKPVNNRPVCTARCGGYGSDGVLSPNYPOVYTSQIC 1404  
 DB 1508 GG-----CGGIIQISKREIHPNYPNNYRAATEC 1536  
 QY 1405 LYFVTVKQVYVFGQAFAFH-TALNDVVEVHDGSHRLLSLSGSHSTESLPLATSNQ 1463  
 DB 1537 SWIIQVRRHHRVLLNTIDPDLAEAPDSCLRLMDSSSTNARVAVACRQGPNNIISGNS 1596

QY 1464 VLIKESAKGLAPARGFHHVQAVPRTSATQCCSVDEPRYGRKLGSDFSVGAIVAFECNCG 1523  
 DB 1597 LFVRSSGSSGNRRFRAEFRE----- 1618  
 QY 1524 YALGSSPEIECLPVGALAAQNNVSAPTCVPCGN-LTERRGITLSPGFPEPLYNLCV 1582  
 DB 1619 -----ECGGIMTDSDTLFSPLXPHNLYHNNCS 1648  
 QY 1583 WKI-VPEGAGIQIOVSF-----VTEQWMSLEVPDGDNTVMLGSGSTTVALLNST 1637  
 DB 1649 WITEQPPPHNITLSTFTLOLONSTDCRDFVELDGNDDAPVQGRYCGSLPHPTISF 1708  
 QY 1638 SNOULHFYSDISVSAGFHELYKTVGLSSCPBPVSNKVGTGERLYVNDVVSFOCEPG 1697  
 DB 1709 GNALTREFTDSTSGFEGRALYSA----- 1733  
 QY 1698 YALGGHAIISCPGTVRRMNPPLCLACCGGTVEMEGVILSPFGPNYPSNMDCSWKI 1757  
 DB 1734 -----STSSCGGSFYTLDDGFNSPDYADYHPNAECVWNI 1768  
 QY 1758 ALPVGFAHIOFLNSTEP-----NHDYIEIRNGPYETSRMWGPPSGSELSLSLT-SHE 1812  
 DB 1769 ASSGNLQSLFSLFENLNSLNCXKDVEIREG-NATGHLIGKICGSLPGNYSAGHS 1827  
 QY 1813 TTVTFHSDHSONRPGFKLEYQ----- 1835  
 DB 1828 LMYRFVSDSGTGSGFQARFNKIFGNNNIVTGHKIASPMPGPKYPSNYSKVVVNDVAY 1887  
 QY 1836 -----ELQECDDPPRANGIVRGVYN-----VG-----QSTVECLP 1868  
 DB 1888 HIHGRLEMDIEFTTTCFDSLKTYGDFHSLRLLGTGCTGTESTSSRNRLTPOFS 1947  
 QY 1869 GYOLTGHPVLTCOHGTRNMNDHPLKCEVPCGNITSSNGTV--YSPGFPSPSSODCV 1926  
 DB 1948 DSSVSGRFLLEMPAVVSDSTPPTIAPGACGGMVTDGPVHIFBPGWREIRANADCI 2007  
 QY 1927 WLTVPFGHVRNLSLQTER-----SGDPTTWGPOQTAPRLGVFTSSMAKTVQSSS 1982  
 DB 2008 WIIYAP-DSTVEINILSLDIEPOOSCNVDLIYDGDSDLSPELAVLCGVSPPGPIRSNG 2066  
 QY 1983 NOVLKFRDATTGATGIFALIFSAVPLTKCP-----PPTILPNAEVTNE 2027  
 DB 2067 EMYVIRTSDTSAVG---TGFMNSFKKSCGGYLAHNRGVITSKPYDTYLPNI-----NC 2118  
 QY 2028 ENNIGDIVRYRCLPGFTL-----VGNELITCKLGYLP-EGP-----PPI 2067  
 DB 2119 SMHV--LVQ-----TGLTIAVHFQOPQIQNRDSFGSGDYLVLRNRPDMSHSPILGSGRN 2172  
 QY 2068 -----CEVHCPT-----NELL-----TDSGTVI 2085  
 DB 2173 GRFCGMVABSTLFTSGNEMFVQFISDSNGGQGFKIRYAKSLACGTVYIHDADSDGL 2232  
 QY 2086 LSQSYGSPYQFQTCMLVREBDYNIISLVEYFLSEKQYDF-----ELPD 2132  
 DB 2233 TSPNYPANYPQHAECWILEAPGRSIOQFE-----DQENIEDTPNCVSYLELD 2284  
 QY 2133 GBSGGSPLLKALSGNSAPLITSSNSYILKMSDPAVNRKGFIRYSAPYCSLPRAPL 2192  
 DB 2285 GANSNRILVSKLGG-HTLPHSWVSRERYTLKFTHGGSSVNGFAKYSIASC----- 2336  
 QY 2193 HGFILOSTQTPGSHFGCNAGYRLVGHSMALCTRHPOGYLMEBALPLCALSCGLPE 2252  
 DB 2337 -----GGTVSGDSGVLE---SIGYPTLPYANNV-----FCQWPIRGLP- 2371  
 QY 2253 APRNGVFGKEYTVGTKAVYSCSEGYHLQAGAEATAECLDTGLMSNR-----NV 2301  
 DB 2372 -----GHYLTLS-----FEDENLQSSPGCTKDPFE--IWNHNSGRVLGYCGNS 2414  
 QY 2302 PPGCVPTVCPDVST-----SVENGRMLLIFETQYQOALMLICDPGYTYTGQVIRIC 2355  
 DB 2415 TBSVD-TSSNVAKYVFTDGSVTASGFLQPKSSRO-----VCG----- 2453

QY 2356 QANGKMSLD-STPTCRITISCELP-----ENG-RI--GLTSLYGATAIF----- 2399  
 DB 2454 -----BDLHGPT-----GTFTSPNYPNPNPHARICEMWITVQEGRRIVLFTNLRL 2499  
 QY 2400 -----SCNSGYTLVGSVRRECMMANGLMSGSEVCLAGHCTPPPIVNGHNGENYSYRGV 2455  
 DB 2500 STQPSGNSBHLIV-----FNGIRSNS-----PLQO----- 2524  
 QY 2456 VYQCNAGFPLIGMSVRIQQO-----DHWSGTPPCVPITTCGHPGNVGLTQGNQFNIN 2510  
 DB 2525 -----KLCRVAVVNEFKSSGNTMKVFFPTDG--SRPYGGL----- 2559  
 QY 2511 DVYKFCVGNWABGAARSQCLASQMSMLPTCIIINCTDGHQBNSTR-----QV 2562  
 DB 2560 -----ASTSTEDAVCGG--FLPSVSGGNFSSPGY--NGIRDYARNLDCW 2601  
 QY 2563 HASGPHRSFGTTVSYRCNHGFFLLGTPLVSCQ-----DGDGTDRDRPQCLLYS 2611  
 DB 2602 TSNPFRNNSISI-----YFLELSIESHQDCTFVLEFRVVDAGPLIEKCSLSA 2653  
 QY 2612 CGHPSPPHSQ-----MGSD--STVGVAVRYS--CTGKRT--LVGNSTRMGLDGM 2658  
 DB 2654 PTAPLVIPPQVWIHFVNSRVEYT-GFYIEYSFTDCGIRTDGNGVILSPVYPLYGAM 2712  
 QY 2659 TGLPHSGSTSVGCGDPIPAHGIRLDSPDPGVMMFSCBAGHVLARGSSERTQANGS 2718  
 DB 2713 T-----HCSMLKAPG-----HTTL-----TFSDFLBA-----HPTDSUV 2747  
 QY 2719 W--SGSQPCGVIS--CGNP-----GTPSNARVFPSSDLVFSSTIVRECGEYVATGLLS 2769  
 DB 2748 TVRNGDSDPSGVIGRYCGQSVPRPIOSGSLIV-----TENTNNOGQTR-GFYA----- 2796  
 QY 2770 RHCSVNGMTSDSPCLVINGDPIRANGRLGNDPFRNKTVYQCYGTMESHRSV 2829  
 DB 2797 -----TWITN-----ALGCGGTFHSANG-----TISPHWPQTFPERSRCSW 2833  
 QY 2830 LSCITD-----RTWNGTKPVCKALMC-----KPPPLIP 2857  
 DB 2834 TWITHESGHWEISPDNSNRIRISDSQCONSVKXWEGRLMINKILLATSCGDVAPSPIVT 2893  
 QY 2858 NGKVVSGDFPMWGSVYTAACLEGYOLSLPAYTCE--GNSWTGELPQCFPVFCGDPG-- 2912  
 DB 2894 SGNL-----FTAVQSEHMAAQGSASFISRCGRFTMTSPDI 2931  
 QY 2913 -VPSRGREDRGFS-----YRSVSFSCHPPLVLVNSPFRFQSDG----- 2952  
 DB 2932 ISPNFPRKQYDNNMCTYLIDADPOSLVILFVSHLEDRSAITGT-----CDHDLHITKG 2987  
 QY 2953 -----TWSGTQ--PSCID-PTLTT-CADPGVPQFIGNNSGCVQSVTLVFRQCK 2998  
 DB 2988 RNLSTPLVITCGSETLRPLTVDDPVLNPFSDAVTTDFGKLSYRAITCGG--IYNSS 3045  
 QY 2999 GYLQSGTTRCLPN 3013  
 DB 3046 GILSPSYSYSNYPN 3060

RESULT 15  
 Q29530 PRELIMINARY; PRT; 2014 AA.  
 AC Q29530.  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
 DE Complement receptor 1 (Fragment).  
 GN Name:CR1.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.  
 OX NCBI\_Taxid=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9432739; PubMed=8021505;

RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;  
 RT "Primary sequence of an alternatively spliced form of CR1. Candidate  
 RT for the 75,000 M(r) complement receptor expressed on chimpanzee  
 erythrocytes."  
 RL J. Immunol. 153:691-700(1994).  
 DR EMBL, L24920; AAAS1438.1; -.  
 DR PIR, I36936; I36936.  
 DR GO, GO:0004872; F:receptor activity; IEA.  
 DR InterPro, IPR000834; Repertoire M4.  
 DR InterPro, IPR001424; SOD CU ZN\_1.  
 DR InterPro, IPR000436; Sush1\_SCR\_CCP.  
 DR Pfam, PF00084; Sush1; 30.  
 DR SMART, SM00032; CCP; 30.  
 DR PROSITE, PS00133; CARBOXYPEPT ZN 2; UNKNOWN\_2.  
 DR PROSITE, PS00087; SOD CU ZN\_1; UNKNOWN\_1.  
 DR PROSITE, PS50923; SUSH1; 30.  
 DR Receptor.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 2014 AA, 221280 MW, 6DC3A74D81F1DB9 CRC64;  
 Query Match 8.2%; Score 1390; DB 2; Length 2014;  
 Best Local Similarity 22.2%; Fred. No. 2.1e-70;  
 Matches 544; Conservative 266; Mismatches 820; Indels 822; Gaps 111;  
 QY CEEDEVPV---YSIRKGLQFGVGDPLTFSCFPYRLGRTARITC-LGRRRLTSSPLPRC 853  
 DB 79 CRNPDPVNGMVAHYIKDIOF--GSOIKSCTKGRILIGSSATCIIISGDVYIMNEPTIC 136  
 QY 854 -VAECGNSVTGTGTLSPNFPVNNHNCIYSIQTOPGKGIQKARAFESGDVLKY 912  
 DB 137 DRICGLPPTITNDGFISTN---RENPHYGSVVTYRCNPGSG--GRKVFEL-VGEP-SI 188  
 QY 913 YDGNNSARKLGVSFHSHEMAGVTLNSTSSSLMDFTDAENTSXGFEIHFSSFEIACED 972  
 DB 189 YCTSNDDQ--VGWNGAPQCIIIPN-----KCTP 215  
 QY 973 PGTAKFGKVADEGHF--AGSSVSFSCDGYSLGSEBELCLSGERRTMDRLPTCYAEC 1030  
 DB 216 PNVEN-GILVSDNSLPSLNEVEFRCPGFVWKPRAVKCOALNK--WEDELSCSRVC 272  
 QY 1031 GGTVRGEVSGQVLSPPGAPAPYEHNLNCIWTIEAENGCTIGLHPLVFTDEVDYLRIMDG 1090  
 DB 273 -----QPP-----DVLH----- 280  
 QY 1091 PVEGVLKELSGPALPKDLHSTNSVVLQFTPTFTSKGFALQFSVSTATSCNDPEIP 1150  
 DB 281 -----A 281  
 QY 1151 QNGSRGDSMEAGSTVQCQDGYALOGSASEICVKIENRFMQSPPTCIAPCGGLTG 1210  
 DB 282 ERTQRDKNFSPGGEVFPSCGYDLRGASLRCTPGD--WSPATPTCVKSCDDPMG 338  
 QY 1211 P--SGVILSPNYPPPYPPKCEKDWKVTVPDYVIALVENIENLEPGYDFLIYDGRDLS 1268  
 DB 339 QLNGRVLFP-----VNLQGLAKVDVDC----- 362  
 QY 1269 PLISFYGSQLPGRIBSSNSLFLA-----FRSDASVSNAGFVIDYTENPRESCFPDPSI 1333  
 DB 363 -----EGFOLKG--SSASYCVLAGMESLWNSVPCVCEQIF-----CPSPPIV 402  
 QY 1334 KNGRVRGSDTLK-----GSSVTYYCH-----GGYVEGTSLSCLGPDGKVMNNPPVC 1374  
 DB 403 PNGNHTGKPLVFPFGKAVNYTCDPHPRGTTFDIIIGESTIRCTSDPGNGWSSPARC 462  
 QY 1375 TAPCGGYVSGDGVLSPNYPONTYTSQICLYFVTPVDYVVFQAFHTALNDVVEVH 1434  
 DB 463 -----GIL----- 465  
 QY 1435 DGHGQ--HSRLLSISGSHTEBSLPATSNOLIKFSAKGLAPARF-----HFVYQAV 1486  
 DB 466 -GHCQAPDHLFLAKLKTQTNASDPFIGTS-----LKYKCRPREYGRFSGITCLDNLWSS- 519

QY 1487 PR--TATQCCSVPEPRYK-RIGSDFSVGAIVREPCNSGYALQSPBEIECLFVPGALQ 1543  
 DB 520 PKDVCKRSCKTTPDPVNGMVAHYITDIOVSRINYSCTGHRILIGHSABCI--LSGAAH 578  
 QY 1544 WNSAPPCV-VPCGGULTERGTTILSPFPEPLNLCWKCWKIVBEGAGIOQVVSFV 1602  
 DB 579 WSTKPPICORIPCGLPPTIANGDFISTRENPHYS-----VVTYRC 620  
 QY 1603 EQWMDSLVEFDADNTVIMLGSFSGTTVPALNLSNOLYHFVSDISVSAAGFHELYKT 1662  
 DB 621 NPSRGRKVE-----LVGPS-----IYCTSD-----DQ 646  
 QY 1663 VGLSCEP--AVPS-----NGVYTER--YLVNDVVSFQCEPGALQHAHISC 1709  
 DB 647 VGWNGAPQCIIIPNCTPPNENGLVSDNRELFSINEVEFRCPGFAMKPRRYKCQ 706  
 QY 1710 PGTVRKNYPPPLCIAOCGGTVEEMGVILSPFGNYPNMDCSKIALPVGFAHIQF 1769  
 DB 707 --ALNKWEPPLPSCSRVC----- 722  
 QY 1770 LNFSTEPNHYIELRNGPYETSRMMGRFSGSELPSLSLSTSHETTVFHSDSQNRPGFK 1829  
 DB 723 -----QPPDVL-----HAETOR----- 736  
 QY 1830 LEQAYELQPCDPPEPPANGIVRAGYNGQSVTFECLPGYOLTGHPVLTQHGTRNMD 1889  
 DB 737 -----DKDN-----SPQEVFYSCEPQYDLRGASLRCT--PGDWS 772  
 QY 1890 HPLPKCEV-PCGNITIS-SNGITYSPGFSPSSQDCWLITVPYIGHGYRLMSLLQTE 1947  
 DB 773 PAAPTECVKSCDFMGOQLNGRVLP-----VNLQGLAKV----- 807  
 QY 1948 PSGDFTIMDQPGOTAPRLGVFTSRMAKTKVQSSNOVLKFRHDAATGIFAIARSAP 2007  
 DB 808 --DFVCD-EGFOLKG-----SSASICVL-----AGMESLWNSVP 839  
 QY 2008 LTK---CPPTIIPNAEVLEN--EENIGDIYVRCLP-----GFTLVGNEIITCKL-- 2055  
 DB 840 VCGQIFCPSPVLPNGHTKPLEVPPFGAVVYTCDPHPRGTTPDLIGESTIRCTSDP 899  
 QY 2056 --GTLYQFEGPPICEV--HCPTNELIDSTGVILISQSYPSGYPOFQTCMWLVREPDY 2110  
 DB 900 QGNGVW--SSPARCGIILGHCOA-----PD- 922  
 QY 2111 NISLTVYELSEKQYDEFELFDGSPGSPILKALSGNYSAPLVTSSNSVYLRMSDHA 2170  
 DB 923 -----HFLFALKTQTNASDPFISLKYECRPEYGRFSTICIDNLV--WSPKD 972  
 QY 2171 YNRKGFIRYSAPYCSLPAPLHGFILOSTQPGSIHFGCAGYRLVGHSMALCTRH 2230  
 DB 973 VCR-----KSKTPPPVNGMVAHYITDIOVSRINYSCTGHRILIGHSABCI--LSG 1024  
 QY 2231 QGYHLMSEALPLCOALSQGLPEAPKNGWVG--KEYTVGKAVYSCSEGYH--LQAG 2283  
 DB 1025 NSAH-WETKPCICORIPCGLPPTIANGDFISTRENPHYSVTVYRCNLGSRGRKVEY 1083  
 QY 2284 AEATACEL---DTGLMS-----NRNVP-----QCPV 2307  
 DB 1084 GEPSTICTSNDDOGVISGAPQCIIIPNCTPPNENGLVSDNRELFSINEVEFRCP 1143  
 QY 2308 -----VTC-----PDVSSIS-----VEHGRRLIFETQYQFOALMLICDP 2343  
 DB 1144 GFWKGRHRYKCOALNMEBELSCSVCVCPPEILHGETTBSHQDNFSGQGVFVSCEP 1203  
 QY 2344 GYVYTGQVRVRCOANGKMSLGDSPTCRILSC--GELPIPNGRHRIGLTV-YGATAI 2398  
 DB 1204 GYDLRGASLHCTPOGDMW--PEAPRCAYVSCDDFLGQL--PHGRILFPLNLQGLAKVS 1258  
 QY 2399 FSCNSGTYLVGSVRCEMANG--LMSGSVRCLAGHCCTPPEIYVNGHNGEYVS--YR 2452  
 DB 1259 FVCDDEGRILGSSVSHCVLVGMRSLWNSVPCVCEQIFCPNPALILNGRHATGTFGDI 1318  
 QY 2453 GSVVYQCN-----AGRLIGMSVRICQDQHH-----WSGKTPTPC-VPTCGH----- 2493

```
Db      1319 KEISYTCDPHPDRGMTFNLIGESTIRCTSDPHGNGVWSSPA PRCELPHVHAGCKTPEQFP 1378
QY      2494 ---PCNPVNLJTGQNOFNLDVVKFVNCNPGYMAEGAARSOCIASGQWMDLPTCRILINCT 2550
Db      1379 PASPTLFINDF---EPVGTSLVNECRPGYFGMFSIS-CLENLVWSSVEDNCRKSCG 1433
QY      2551 DPGHOENSVRQVHASGPHRFSFGTTVSYRCNHGFYLLGTPLVSCQGDG---TWRDRPQC 2607
Db      1434 PPEPFNGVWHINTD---TQFGSTVYSCNEGRLLGSPSTCIVSGNNVTWDXKAPIC 1489
QY      2608 LNVSCGHPSPPHSQMSGDSYT-----VGAVRYS-----IGKRLVGN 2648
Db      1490 EIIISC---EPPTISNGDFYSNNRASFNHGTVTYQCHTGPDEQLFELVGERSIY-C 1543
QY      2649 TRMCGLDGHTGSLPHCSGTSVGVCGDPCI PAHGIRLGDS---FDPGTWRFSCGAVL 2705
Db      1544 TSDDDGVWSSPPRCLISTN--KCTAPEV-ENAIRVPGNRSPFSLTEIVFRCPQGFVA 1600
QY      2706 RGSSERTQANGWSGSOPECGVISCNP-----GTPSNARVPSDGLVFSSSIYEC 2758
Db      1601 VGSHTVQCCQTNMGWMPRLPHCSR-V-CQPPPEILHGEHTPSH-----QDNFSPGQEVFYGC 1654
QY      2759 REGYVATGLLSRHC SVNGTGTSGDPCLVINCBD--PGIPANGRLGNDFRNKVITYQC 2816
Db      1655 EPGYDLRGAASLHCTPQDWSPEAPRCTVXSCDDFLQPLHGRVLPFLNLQLGAKVSFVC 1714
QY      2817 VPGYMMESHRSV--VLSCTDRTMNGTKPYCKALMCKPPLIPNGKVGS---DFMWS 2871
Db      1715 DEGFRLKGRASHCILGKMK-ALMNSVFPVCEQIFCENPPAILNGRHGTGTPFGDIPYKE 1773
QY      2872 VTYAC-----LEGYQLSLPAVFTCEGNGSWTGLPQC--FPVFCGDPGVPSRG 2917
Db      1774 ISYACDTHPRGMTFNLIG-ESSIRCTSDPQNGVWSSPA PRCELSVPAACPHP--PKIQ 1830
QY      2918 RREDRG-----FSYRSSVSFSCHPPLVYVGSPPRFCSQSDGTSCTOPSCIDPTLTTCADP 2972
Db      1831 NGHDIGHVSILYLPGMTISYICDPGYLLVGKGFIFCTDGIWSQLDHYCKE--VNCSP 1887
QY      2973 GVPQFGIOWN---SOGYVQSTVLFRCQKGYLLQGSTTRTCLPNLTWSGTTP 3021
Db      1888 -LFMNGISKELEMKVYHYGDYVTLKCEDGYTLEGSFWSQCADRMD--PP 1936
```

Search completed: October 18, 2004, 10:24:50  
Job time : 191 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 09:57:10 ; Search time 109 Seconds  
(without alignments)  
10215.557 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985

Sequence: 1 MACAPPALLPCLSLSDCC.....RSGPVDPSTLPESHSPKR 3104

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 35872929 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_23Sep04:.\*  
1: geneseqp1980s:.\*  
2: geneseqp1980s:.\*  
3: geneseqp2000s:.\*  
4: geneseqp2001s:.\*  
5: geneseqp2002s:.\*  
6: geneseqp2003as:.\*  
7: geneseqp2003bs:.\*  
8: geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 16985 | 100.0       | 3104   | 5  | ABG79168 Human cub |
| 2          | 16985 | 100.0       | 3104   | 8  | ADH71166 Human pro |
| 3          | 16985 | 100.0       | 3104   | 8  | ADH71138 Human pro |
| 4          | 16978 | 100.0       | 3104   | 8  | ADH71168 Human pro |
| 5          | 16057 | 94.5        | 3546   | 8  | ADH71136 Human pro |
| 6          | 15655 | 92.2        | 3483   | 8  | ADH71144 Human pro |
| 7          | 15219 | 89.6        | 3130   | 8  | ADH71146 Human pro |
| 8          | 14142 | 83.3        | 2669   | 8  | ABG79169 Human pro |
| 9          | 14142 | 83.3        | 2669   | 8  | ADH71140 Human pro |
| 10         | 13142 | 77.4        | 2612   | 8  | ADH71142 Human pro |
| 11         | 11247 | 66.2        | 3567   | 8  | ADH72216 Human pro |
| 12         | 10760 | 63.3        | 3100   | 5  | AAE20789 Human C3b |
| 13         | 10755 | 63.3        | 3100   | 5  | AAE20901 Human C3b |
| 14         | 10720 | 63.1        | 3095   | 5  | AAE20788 Human C3b |
| 15         | 10701 | 63.0        | 3069   | 5  | AAE20787 Human C3b |
| 16         | 10696 | 63.0        | 3069   | 5  | AAE20900 Human C3b |
| 17         | 10475 | 61.7        | 3389   | 7  | ADJ70480 Human hea |
| 18         | 9025  | 53.1        | 1958   | 5  | AAO19415 Human mol |
| 19         | 7662  | 45.0        | 2352   | 5  | AAU11816 Human can |
| 20         | 7643  | 45.0        | 2306   | 5  | AAU11817 Human can |
| 21         | 7568  | 44.6        | 2008   | 5  | AAU11814 Human can |
| 22         | 7098  | 41.8        | 1826   | 5  | AAU11812 Human can |
| 23         | 6949  | 40.9        | 1800   | 5  | AAU11813 Human can |
| 24         | 6936  | 40.8        | 1783   | 5  | AAU11815 Human can |
| 25         | 4937  | 29.1        | 882    | 4  | ABG83372 Human pro |

# ALIGNMENTS

|    |      |      |      |   |                     |
|----|------|------|------|---|---------------------|
| 26 | 4937 | 29.1 | 882  | 8 | ADH71164 Human pro  |
| 27 | 4869 | 28.7 | 884  | 4 | ABG83371 Human pro  |
| 28 | 4859 | 28.6 | 884  | 8 | ADH71162 Human pro  |
| 29 | 4692 | 27.6 | 869  | 8 | ADH71154 Human pro  |
| 30 | 4268 | 25.1 | 776  | 8 | ADH71150 Human pro  |
| 31 | 4257 | 25.1 | 776  | 8 | ADH71148 Human pro  |
| 32 | 4143 | 24.4 | 1274 | 5 | ABG76507 Human enco |
| 33 | 3856 | 22.7 | 713  | 8 | ADH71152 Human pro  |
| 34 | 3174 | 18.7 | 613  | 4 | AAU00816 Human imm  |
| 35 | 2979 | 17.5 | 557  | 8 | ADH71158 Human pro  |
| 36 | 2651 | 15.6 | 810  | 8 | ADH72220 Human pro  |
| 37 | 2639 | 15.5 | 529  | 5 | ABP43478 Human sec  |
| 38 | 2555 | 15.0 | 839  | 5 | AAE33384 Human int  |
| 39 | 2416 | 14.2 | 601  | 7 | ADG31384 Human nov  |
| 40 | 2105 | 12.4 | 413  | 7 | ADG31488 Human nov  |
| 41 | 1850 | 10.9 | 3594 | 5 | AAE20147 Human C3b  |
| 42 | 1794 | 10.6 | 3571 | 5 | AAE20146 Human C3b  |
| 43 | 1793 | 10.6 | 3570 | 5 | ABJ10589 Human nov  |
| 44 | 1793 | 10.6 | 3570 | 8 | ADH71302 Human pro  |
| 45 | 1793 | 10.6 | 3570 | 8 | ADH71304 Human pro  |

# RESULT 1

ABG79168 standard; protein, 3104 AA.

ABG79168; (first entry)

15-NOV-2002 (first entry)

Human cub and sushi domain containing protein #1.

KW Human; NOVX; neurodegenerative disease; Alzheimer's disease; anxiety;  
KW Parkinson's disease; Huntington's disease; neurological disorder;  
KW schizophrenia; manic depression; mental retardation; angina pectoris;  
KW cardiovascular disease; acute heart failure; myocardial infarction;  
KW muscular disease; muscular disorder; retinal disease; photoreception;  
KW deafness; keratinisation disorder; cancer; ovarian cancer; melanoma;  
KW immunological disorder; inflammatory disease; immune disease; diabetes;  
KW bacterial infection; fungal infection; protozoal infection; obesity;  
KW viral infection; reproductive system disorder; metabolic disturbance;  
KW anorexia; wasting disorder; chronic disease; infectious disease;  
KW dyslipidaemia; cub; sushi; myelin; von Willebrand factor; kielin;  
KW semaphorin; serine/threonine protein kinase; TGF-beta binding;  
KW mas proto-oncogene; ribonuclease pancreatic precursor; aminotransferase;  
KW toll-like 2; cysteine sulfinic acid decarboxylase SNP;  
KW single nucleotide polymorphism.

Homo sapiens.

Key Location/Qualifiers

FT misc\_difference 1027 /note="Ala substituted by Thr as a result of a single nucleotide polymorphism (SNP)"

FT misc\_difference 1084 /note="Val substituted by Ala as a result of a single nucleotide polymorphism (SNP)"

FT misc\_difference 1362 /note="Asp substituted by Gly as a result of a single nucleotide polymorphism (SNP)"

FT WO200264791-A2.

FT 22-AUG-2002.

FT 10-DEC-2001; 2001WO-US046369.

FT 08-DEC-2000; 2000US-0254329P.

FT 14-DEC-2000; 2000US-0255648P.

FT 15-MAY-2001; 2001US-0291037P.

PR 08-JUN-2001; 2001US-0297173P.  
 PR 08-JUN-2001; 2001US-0309258P.  
 PR 28-AUG-2001; 2001US-0315639P.  
 PR 01-OCT-2001; 2001US-0326393P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Alabrook JP, Anderson DM, Burgess CE, Boldog FL, Casman ST,  
 PI Colman SD, Ringer SR, Ellerman K, Gerlach V, Gorman L, Grose WM,  
 PI Guo X, Hartmann JL, Kekuda R, Lepley DM, Li L, Macdougall JR,  
 PI Miller I, Pena CBA, Peyman JA, Rastelli L, Rieger DK, Shinkels RA,  
 PI Smitheon G, Spytek KA, Stone DJ, Tchernev VT, Vernet CAM, Voss EZ,  
 PI Zetunuen BD, Zhong H, Zhong M;  
 XX  
 DR WPI; 2002-643486/69.  
 DR N-PSDB; ABS64375.  
 XX  
 PT New NOVX polypeptides and polynucleotides useful for treating or  
 PT preventing e.g. neurodegenerative diseases, neurological disorders,  
 PT cardiovascular diseases, muscular diseases and disorders, or  
 PT immunological diseases.  
 XX  
 PS Claim 1; Page 13-14; 239pp; English.  
 XX  
 CC The present invention relates to new NOVX polypeptides. The polypeptides,  
 CC polynucleotides and antibodies are useful in the manufacture of a  
 CC medicament for treating or preventing neurodegenerative diseases (e.g.  
 CC Alzheimer's disease, Parkinson's disease, or Huntington's disease),  
 CC neurological disorders (e.g. anxiety, schizophrenia, manic depression or  
 CC mental retardation), cardiovascular disease (e.g. acute heart failure,  
 CC angina pectoris or myocardial infarction), muscular diseases and  
 CC disorders, retinal diseases (including those involving photoreception,  
 CC deafness and keratinisation disorders), cancer (e.g. ovarian cancer or  
 CC melanoma), immunological disorders, inflammatory and immune diseases,  
 CC bacterial, fungal, protozoal and viral infections, and reproductive  
 CC system disorders. The proteins of the invention may be used to screen  
 CC drugs or compounds that modulate the NOVX protein activity or expression,  
 CC as well as to treat disorders characterised by insufficient or excessive  
 CC production of NOVX protein or protein forms that have decreased or  
 CC aberrant activity compared to NOVX wild type protein, such as diabetes,  
 CC obesity, metabolic disturbances associated with obesity, anorexia and  
 CC wasting disorders associated with chronic diseases and various cancers,  
 CC infectious diseases and various dyslipidaemias. The nucleic acid  
 CC sequences of the invention may be used in chromosome mapping, identifying  
 CC an individual from minute biological samples (tissue typing), and in  
 CC forensic identification of a biological sample. The present amino acid  
 CC sequence represents a NOVX protein of the invention  
 XX  
 SQ Sequence 3104 AA;  
 Query Match 100.0%; Score 16985; DB 5; Length 3104;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAGAPPALLPCLSLSDCCASNQRHSVGVSELYKKOIELKSRVKLMPKSDMSQKTS 60  
 DB 1 MAGAPPALLPCLSLSDCCASNQRHSVGVSELYKKOIELKSRVKLMPKSDMSQKTS 60  
 QY 61 VLTQVGVSGHNNCPDPIPERGKRLGSDFRLGSSVQFTCNBGYDQSKRTCKMVSMD 120  
 DB 61 VLTQVGVSGHNNCPDPIPERGKRLGSDFRLGSSVQFTCNBGYDQSKRTCKMVSMD 120  
 QY 121 FAAMSHRPVPCARMCDALHGRSGIITSPNFIQYDNNAHCWITITLNPSTKILAFE 180  
 DB 121 FAAMSHRPVPCARMCDALHGRSGIITSPNFIQYDNNAHCWITITLNPSTKILAFE 180  
 QY 121 EFDLEGGYDITLVGGGGDGDOKTVLYMSQNA CSDBPHTPGSRIPESNSGDIWRQKTVL 240  
 DB 121 EFDLEGGYDITLVGGGGDGDOKTVLYMSQNA CSDBPHTPGSRIPESNSGDIWRQKTVL 240  
 QY 181 EFDLEGGYDITLVGGGGDGDOKTVLYMSQNA CSDBPHTPGSRIPESNSGDIWRQKTVL 240  
 DB 181 EFDLEGGYDITLVGGGGDGDOKTVLYMSQNA CSDBPHTPGSRIPESNSGDIWRQKTVL 240  
 QY 241 EICRDISSSDAAGSVRKSPKTSNAVELVAPGTLEIGSGCGDPGIPAYGRREGSRFHND 300  
 DB 241 EICRDISSSDAAGSVRKSPKTSNAVELVAPGTLEIGSGCGDPGIPAYGRREGSRFHND 300  
 QY 241 EICRDISSSDAAGSVRKSPKTSNAVELVAPGTLEIGSGCGDPGIPAYGRREGSRFHND 300  
 DB 241 EICRDISSSDAAGSVRKSPKTSNAVELVAPGTLEIGSGCGDPGIPAYGRREGSRFHND 300

QY 301 TLKECOPAFELVGVQKAITCOKNNQWAKKPGCVSCFPNFTSPSGVLSPNVPEDYGNH 360  
 DB 301 TLKECOPAFELVGVQKAITCOKNNQWAKKPGCVSCFPNFTSPSGVLSPNVPEDYGNH 360  
 QY 361 LHCWVLLARPESRHLAFNDIDVEPQFVLTKGATAEAVLGTFSGNOLPSSITSSG 420  
 DB 361 LHCWVLLARPESRHLAFNDIDVEPQFVLTKGATAEAVLGTFSGNOLPSSITSSG 420  
 QY 421 HVARLEFQDHS TGKRGFNITFTFRHNECPDPGVVNGKRGDSIQGLSSISFLCDSEGF 480  
 DB 421 HVARLEFQDHS TGKRGFNITFTFRHNECPDPGVVNGKRGDSIQGLSSISFLCDSEGF 480  
 QY 481 LGTQSEITTCVLKRGSVVNSAVLRCEAPCGHLTPSBGTTLSBPMGPYKDALSCAWV 540  
 DB 481 LGTQSEITTCVLKRGSVVNSAVLRCEAPCGHLTPSBGTTLSBPMGPYKDALSCAWV 540  
 QY 541 IEAOGYPIKIFDEBFKTEVNDLEVRDGRFYSAPLIGVYHGNVQPOFLITSNYVL 600  
 DB 541 IEAOGYPIKIFDEBFKTEVNDLEVRDGRFYSAPLIGVYHGNVQPOFLITSNYVL 600  
 QY 601 FSTDKSHSDIGFQLRYETITLQSDHCLDPGIPVNGQRHNDPFYVGLATVFSGDSGYTSD 660  
 DB 601 FSTDKSHSDIGFQLRYETITLQSDHCLDPGIPVNGQRHNDPFYVGLATVFSGDSGYTSD 660  
 QY 661 GEPLECEPFWMSRALPSCCALCGFICGSSGTTLSPEPDPYNNLCTWIIETSHSGK 720  
 DB 661 GEPLECEPFWMSRALPSCCALCGFICGSSGTTLSPEPDPYNNLCTWIIETSHSGK 720  
 QY 721 VFFTHTHLBSGHDYLLITENGSTFQPLRLQUTGSRLLPAPISAGLYGNTQVRIQSF 780  
 DB 721 VFFTHTHLBSGHDYLLITENGSTFQPLRLQUTGSRLLPAPISAGLYGNTQVRIQSF 780  
 QY 781 MSYEFNITFSEYDLEPCEBEVPAVSIKRLQGVGDTLTFSCPGYRLGTARITLG 840  
 DB 781 MSYEFNITFSEYDLEPCEBEVPAVSIKRLQGVGDTLTFSCPGYRLGTARITLG 840  
 QY 841 GRRRLMSPLRVCVAECNSVTGQTLISNPNVNNNECTYISITQGGKIQQLAR 900  
 DB 841 GRRRLMSPLRVCVAECNSVTGQTLISNPNVNNNECTYISITQGGKIQQLAR 900  
 QY 901 AFELSEGVDLVKYDGNNSARLLGVFSHEMVGVTLSSTSSLWLDFTDAENTSKGEL 960  
 DB 901 AFELSEGVDLVKYDGNNSARLLGVFSHEMVGVTLSSTSSLWLDFTDAENTSKGEL 960  
 QY 961 HFSSELIKCDPCTPRKGVYVHDEGHPAGSVSFCDDPGYSLGSEBELCLSGERTWD 1020  
 DB 961 HFSSELIKCDPCTPRKGVYVHDEGHPAGSVSFCDDPGYSLGSEBELCLSGERTWD 1020  
 QY 1021 RPLPTVAECGTVRGEVSGVLSPGYPAPYEHNLNCTWTEBAGCTIGLHFLVPTDEE 1080  
 DB 1021 RPLPTVAECGTVRGEVSGVLSPGYPAPYEHNLNCTWTEBAGCTIGLHFLVPTDEE 1080  
 QY 1081 VHDVLRITWDGVESGVLLKELSGPALPKDLHSTFNSVLQFSTDFPSKQGFALQFSVST 1140  
 DB 1081 VHDVLRITWDGVESGVLLKELSGPALPKDLHSTFNSVLQFSTDFPSKQGFALQFSVST 1140  
 QY 1141 ATSCNDPGIIPONGRSRSDSWEAGDSTVFQCDPGIALQGSABEISVKNRFPQPSPTC 1200  
 DB 1141 ATSCNDPGIIPONGRSRSDSWEAGDSTVFQCDPGIALQGSABEISVKNRFPQPSPTC 1200  
 QY 1201 IAPCGDULTPGSGVLSPNVPEYPPGKECDMWKTVSPDVIVLALFVNIENRQYDPLHI 1260  
 DB 1201 IAPCGDULTPGSGVLSPNVPEYPPGKECDMWKTVSPDVIVLALFVNIENRQYDPLHI 1260  
 QY 1261 YDGRDLSPLIGSYGSLQGRLESSNSGLFLAFRSDASVSNAGFVIDYENRRESCFDP 1320  
 DB 1261 YDGRDLSPLIGSYGSLQGRLESSNSGLFLAFRSDASVSNAGFVIDYENRRESCFDP 1320  
 QY 1321 GSIKNGTRVSDKLKSSVTVYCHGGYEVGTSLLSCILGPDGKPVWNNRPVCTAPCGG 1380  
 DB 1321 GSIKNGTRVSDKLKSSVTVYCHGGYEVGTSLLSCILGPDGKPVWNNRPVCTAPCGG 1380

QY 1381 QYVSGDVVLSPNYPONTSGOICLYEYVTPKDYVVGQAFPAFTALNDVVEVHDGSH 1440  
 DB 1381 QYVSGDVVLSPNYPONTSGOICLYEYVTPKDYVVGQAFPAFTALNDVVEVHDGSH 1440  
 QY 1441 SRLLSLSGSHTESEPLATSNQVLKFSAKGLAPAGFHFVYQAVPRISATOCSSVPE 1500  
 DB 1441 SRLLSLSGSHTESEPLATSNQVLKFSAKGLAPAGFHFVYQAVPRISATOCSSVPE 1500  
 QY 1501 RYKRLASDSVGAIVAFECNSGVALOGSPBIECLPVPGLAOMNVSAPFCVPCGNLT 1560  
 DB 1501 RYKRLASDSVGAIVAFECNSGVALOGSPBIECLPVPGLAOMNVSAPFCVPCGNLT 1560  
 QY 1561 ERGGTILSPGPPEPYLNSLCWKIVYBEGAGIOIVVSVTEBOMNDSLEVPDQADTVT 1620  
 DB 1561 ERGGTILSPGPPEPYLNSLCWKIVYBEGAGIOIVVSVTEBOMNDSLEVPDQADTVT 1620  
 QY 1621 MGSFSGTTPVALLNSTNSQLYLFYSDSVSAAGFLYKTVGLSSCPBPVPSNGVKT 1680  
 DB 1621 MGSFSGTTPVALLNSTNSQLYLFYSDSVSAAGFLYKTVGLSSCPBPVPSNGVKT 1680  
 QY 1681 GERLYVNDVVSFOCEPGYALOGNAHISCMGTVRMNYPPPLCIAOCGTVBEGVYLS 1740  
 DB 1681 GERLYVNDVVSFOCEPGYALOGNAHISCMGTVRMNYPPPLCIAOCGTVBEGVYLS 1740  
 QY 1741 PGFPGNVPMSMDSMKIALPVGGAHIOPLNFSTEPHDIETIRNGPYETSRMMGRPSGS 1800  
 DB 1741 PGFPGNVPMSMDSMKIALPVGGAHIOPLNFSTEPHDIETIRNGPYETSRMMGRPSGS 1800  
 QY 1801 ELPSLSLSHSTETTVVYFHSDSHSONRPGFKLEYQAYELOECRDEBPFPANGLVRGAGVNG 1860  
 DB 1801 ELPSLSLSHSTETTVVYFHSDSHSONRPGFKLEYQAYELOECRDEBPFPANGLVRGAGVNG 1860  
 QY 1861 SVTFECLPGYQLTGHPLVTCQHGTRNMDHPLKCEVPCCGNITSSNGITVSRGFPSPYS 1920  
 DB 1861 SVTFECLPGYQLTGHPLVTCQHGTRNMDHPLKCEVPCCGNITSSNGITVSRGFPSPYS 1920  
 QY 1921 SSODCWLITVPYIGHGRLNLSLQTEPSGDFITIMGPQOATPRLGVPFRSMKKTVO 1980  
 DB 1921 SSODCWLITVPYIGHGRLNLSLQTEPSGDFITIMGPQOATPRLGVPFRSMKKTVO 1980  
 QY 1981 SSNOVLKFKHRDADATGCIFAIAFSAYBLTKCPPTILPNAEVTENEENFNGIDIVYRCL 2040  
 DB 1981 SSNOVLKFKHRDADATGCIFAIAFSAYBLTKCPPTILPNAEVTENEENFNGIDIVYRCL 2040  
 QY 2041 PGFTLVGNELITCKLGTLYLOFEGPPICEVHCPTNELLTOSTGYVLSQSPGSPQOTC 2100  
 DB 2041 PGFTLVGNELITCKLGTLYLOFEGPPICEVHCPTNELLTOSTGYVLSQSPGSPQOTC 2100  
 QY 2101 SWIVRVEPDYNISETVEYFLSEKQYDFEIPDGSQSPILKALSGNYSAPLIVTSSNS 2160  
 DB 2101 SWIVRVEPDYNISETVEYFLSEKQYDFEIPDGSQSPILKALSGNYSAPLIVTSSNS 2160  
 QY 2161 VYLWMSDHAUNRGFKIRYSAPYCSLPRAPLHFIIGOSTPGGSIHFGCNAAGYLVG 2220  
 DB 2161 VYLWMSDHAUNRGFKIRYSAPYCSLPRAPLHFIIGOSTPGGSIHFGCNAAGYLVG 2220  
 QY 2221 HSNMICTRHPQYHLMSEAIPLCOALSGLPEADKCNMVFGEKXYTGTAKVYSCSEGYH 2280  
 DB 2221 HSNMICTRHPQYHLMSEAIPLCOALSGLPEADKCNMVFGEKXYTGTAKVYSCSEGYH 2280  
 QY 2281 OAGAEAAECLDTCGLMSNRNVPQCVPTCPDVSSIVEHGRMLIETQYQFOAOLMLI 2340  
 DB 2281 OAGAEAAECLDTCGLMSNRNVPQCVPTCPDVSSIVEHGRMLIETQYQFOAOLMLI 2340  
 QY 2341 CDBPYVYTGQVIRICQANGKMSLDSTPTCHIIISCEGLPIPNHNRIGTISVYGATAIFS 2400  
 DB 2341 CDBPYVYTGQVIRICQANGKMSLDSTPTCHIIISCEGLPIPNHNRIGTISVYGATAIFS 2400  
 QY 2401 CNSGYTLLVGRVRECMANGMSGSEVRCLAGHCTPPEPIYNGHINGENYSTRGSVVYQCN 2460  
 DB 2401 CNSGYTLLVGRVRECMANGMSGSEVRCLAGHCTPPEPIYNGHINGENYSTRGSVVYQCN 2460  
 QY 2461 AGFRLIGMSVRLICQODHMSGKTPFCVPLTCGHGNGVNLGTQONQNLNDVYVFCVNGP 2520

DB 2461 AGFRLIGMSVRLICQODHMSGKTPFCVPLTCGHGNGVNLGTQONQNLNDVYVFCVNGP 2520  
 QY 2521 YNABGAARQCLASGWSMDLPTCRLIINCTDEHONSROVHASGPHRPSFCTTYSYC 2580  
 DB 2521 YNABGAARQCLASGWSMDLPTCRLIINCTDEHONSROVHASGPHRPSFCTTYSYC 2580  
 QY 2581 NHGFYLLGTPLYVSCQDGTMDRPPQCLVSCGHPSPPHOSMGSDSYTGAIVRYSCTG 2640  
 DB 2581 NHGFYLLGTPLYVSCQDGTMDRPPQCLVSCGHPSPPHOSMGSDSYTGAIVRYSCTG 2640  
 QY 2641 KRTLGNSTRMCLDHWMTGSLPHCSGTSVGCGDPIPAHIGIRLDSFDPGTVMRFSCE 2700  
 DB 2641 KRTLGNSTRMCLDHWMTGSLPHCSGTSVGCGDPIPAHIGIRLDSFDPGTVMRFSCE 2700  
 QY 2701 AGHVLKSSBRTQANGSWSGSGOPECGVISCNPGTSPNARVVSGLVYFSSIVYECRE 2760  
 DB 2701 AGHVLKSSBRTQANGSWSGSGOPECGVISCNPGTSPNARVVSGLVYFSSIVYECRE 2760  
 QY 2761 GYATGLLSRHGVSNGTMTGSDPECLVINGDPIRANGRLGNDPRYKTVTYOCVPGY 2820  
 DB 2761 GYATGLLSRHGVSNGTMTGSDPECLVINGDPIRANGRLGNDPRYKTVTYOCVPGY 2820  
 QY 2821 MESHRSVYLSCTKDRTNMGTKRVCALMKCPPLIPNGKVSGSDPMWGSVYACLEGY 2880  
 DB 2821 MESHRSVYLSCTKDRTNMGTKRVCALMKCPPLIPNGKVSGSDPMWGSVYACLEGY 2880  
 QY 2881 QLSLPAVFTCEGNGSWTGBELPOCFPVPCGDPVPSGRNEDRCPSTRSSVSEFCRPLVL 2940  
 DB 2881 QLSLPAVFTCEGNGSWTGBELPOCFPVPCGDPVPSGRNEDRCPSTRSSVSEFCRPLVL 2940  
 QY 2941 VGSRRFCOSDGTWSTGSDCIDLPTLTTCADPVPFGIOMNSOGYOVSTVLPFOCKY 3000  
 DB 2941 VGSRRFCOSDGTWSTGSDCIDLPTLTTCADPVPFGIOMNSOGYOVSTVLPFOCKY 3000  
 QY 3001 LLOGSTTRCTPLNLTWSTGTPPDCVPHNCRQPETPTANVAGALDLPBMGYTLTPPARASP 3060  
 DB 3001 LLOGSTTRCTPLNLTWSTGTPPDCVPHNCRQPETPTANVAGALDLPBMGYTLTPPARASP 3060  
 QY 3061 SRVAPSTAPARRMAAOASBPASMRSGPVGDPSLTGSHRSPKP 3104  
 DB 3061 SRVAPSTAPARRMAAOASBPASMRSGPVGDPSLTGSHRSPKP 3104  
 RESULT 2  
 ADH71166 standard; protein; 3104 AA.  
 ADH71166:  
 25-MAR-2004 (first entry)  
 Human protein of the invention NOVap SEQ ID NO:62.  
 human; cytosolic; immunomodulator; neuroprotective; nootropic;  
 anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;  
 vacine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;  
 obesity; diabetes; infectious disease; metabolic syndrome X;  
 dyslipidemia.  
 Homo sapiens.  
 WO2003102155-A2.  
 11-DEC-2003.  
 03-JUN-2003; 2003MO-US017430.  
 03-JUN-2002; 2002US-0385120P.  
 04-JUN-2002; 2002US-0385784P.  
 05-JUN-2002; 2002US-0386041P.  
 05-JUN-2002; 2002US-0386047P.  
 06-JUN-2002; 2002US-0386376P.

[illegible]

QY 481 LGTQSGSETTTCULKEGSVVWNSAVLRCEAPCGHLTSPSGTILSPGPGFYKDALSCAW 540  
 DB 481 LGTQSGSETTTCULKEGSVVWNSAVLRCEAPCGHLTSPSGTILSPGPGFYKDALSCAW 540  
 QY 541 IEAOPGPKITITDRFKTEVNVDTLEVRDRTYSAPLIGVHGTOVPOPLISTSNLYLL 600  
 DB 541 IEAOPGPKITITDRFKTEVNVDTLEVRDRTYSAPLIGVHGTOVPOPLISTSNLYLL 600  
 QY 601 PSTDKSHSDIGFOLRYETITLQSDHCLDPGI PVNGOHGNDPYVALVTFSCDSGYTLSD 660  
 DB 601 PSTDKSHSDIGFOLRYETITLQSDHCLDPGI PVNGOHGNDPYVALVTFSCDSGYTLSD 660  
 QY 661 FSTDKSHSDIGFOLRYETITLQSDHCLDPGI PVNGOHGNDPYVALVTFSCDSGYTLSD 720  
 DB 661 FSTDKSHSDIGFOLRYETITLQSDHCLDPGI PVNGOHGNDPYVALVTFSCDSGYTLSD 720  
 QY 721 VFTFTFTHLESCHDYLLITENGSTFOTPLQTLGSRLPAPISAGLYGNFTAQVRFISDFS 780  
 DB 721 VFTFTFTHLESCHDYLLITENGSTFOTPLQTLGSRLPAPISAGLYGNFTAQVRFISDFS 780  
 QY 781 MSYEGFNITSEYDLEBCEEPVAYSIRKQLOFVGDTLTFSCFPGYRLEGTAITCLG 840  
 DB 781 MSYEGFNITSEYDLEBCEEPVAYSIRKQLOFVGDTLTFSCFPGYRLEGTAITCLG 840  
 QY 841 GRRRLMSSPLPRCAVECGNSVTGTOCTLLSPNFVANNHNECTYSTOTQPGKIOUKAR 900  
 DB 841 GRRRLMSSPLPRCAVECGNSVTGTOCTLLSPNFVANNHNECTYSTOTQPGKIOUKAR 900  
 QY 901 AFELSEGDVLYKVDGNNSARLLGVFSEMMGVTLNSTSSLMDELDTAENTSKGFE 960  
 DB 901 AFELSEGDVLYKVDGNNSARLLGVFSEMMGVTLNSTSSLMDELDTAENTSKGFE 960  
 QY 961 HFSSFELIKCEDPCTPKFGYKVDHDEGHFAGSSVSFSCDPGYSLRGSEELLCLSGERTWD 1020  
 DB 961 HFSSFELIKCEDPCTPKFGYKVDHDEGHFAGSSVSFSCDPGYSLRGSEELLCLSGERTWD 1020  
 QY 1021 RPLPTCAECCGTVRGEVSGQVLSPGYPARYENHLMCIWTEAEAGCTIGLHLVPEETEE 1080  
 DB 1021 RPLPTCAECCGTVRGEVSGQVLSPGYPARYENHLMCIWTEAEAGCTIGLHLVPEETEE 1080  
 QY 1081 VHVULRIWDPVRESGVLLKELSGPALPKDLHSTFNSVVLQFSTDFTSKOGFALOFSVST 1140  
 DB 1081 VHVULRIWDPVRESGVLLKELSGPALPKDLHSTFNSVVLQFSTDFTSKOGFALOFSVST 1140  
 QY 1141 ATSCNDPGLPONGSRGSDSWAEGDSTVFQCDPGYALQGSABISCVKIENRFFMOPSPPTC 1200  
 DB 1141 ATSCNDPGLPONGSRGSDSWAEGDSTVFQCDPGYALQGSABISCVKIENRFFMOPSPPTC 1200  
 QY 1201 IAPCGGDLTGPSCVYILSPNPEPYPKCECDMKYTVSPDYVALVFNILPEBGPDLAT 1260  
 DB 1201 IAPCGGDLTGPSCVYILSPNPEPYPKCECDMKYTVSPDYVALVFNILPEBGPDLAT 1260  
 QY 1261 YDPRDLSPLIGSFYGSQLPGRJESSNSLFLAFRSPASVSNAGFVLDYENPRESCFDP 1320  
 DB 1261 YDPRDLSPLIGSFYGSQLPGRJESSNSLFLAFRSPASVSNAGFVLDYENPRESCFDP 1320  
 QY 1321 GSIKNGTRVGSDLKLGSSVTTYCHGYEVEGTSTLSCILGPDGKPVNNRPRVCTACGG 1380  
 DB 1321 GSIKNGTRVGSDLKLGSSVTTYCHGYEVEGTSTLSCILGPDGKPVNNRPRVCTACGG 1380  
 QY 1381 QYVGSDDVULSPNPNYTSGOICLYVTVPKDVVGQPAFFFTALNDVVEVDHGSOH 1440  
 DB 1381 QYVGSDDVULSPNPNYTSGOICLYVTVPKDVVGQPAFFFTALNDVVEVDHGSOH 1440  
 QY 1441 SRLLSLSGSHGTESLPLATSNQVLKFSKAGLAPAGFHVQAVERTSATQCSVPEP 1500  
 DB 1441 SRLLSLSGSHGTESLPLATSNQVLKFSKAGLAPAGFHVQAVERTSATQCSVPEP 1500  
 QY 1501 RYGRKLGSDFSVAIVAFECNSGYALQGSPEIECLPVPGALAQNNVAPFCVVPCCGNTL 1560  
 DB 1501 RYGRKLGSDFSVAIVAFECNSGYALQGSPEIECLPVPGALAQNNVAPFCVVPCCGNTL 1560

QY 1561 ERRGTLSPGPEPYLNSLNCVWKIVPEGAGIQIOVSEFTEQNMDSLEVPDADNTVT 1620  
 DB 1561 ERRGTLSPGPEPYLNSLNCVWKIVPEGAGIQIOVSEFTEQNMDSLEVPDADNTVT 1620  
 QY 1621 MGSFSGTIVPALLNSTSNQYLHAFYSDSVSAAGFHELYKTVGLSSCEBPAPVNGVKT 1680  
 DB 1621 MGSFSGTIVPALLNSTSNQYLHAFYSDSVSAAGFHELYKTVGLSSCEBPAPVNGVKT 1680  
 QY 1681 GERLYVNDVVSFCCEGYALQGHANHSCMPGYTRANNYPPLCIACCGGTVEBMEGVILS 1740  
 DB 1681 GERLYVNDVVSFCCEGYALQGHANHSCMPGYTRANNYPPLCIACCGGTVEBMEGVILS 1740  
 QY 1741 PGFPNGYPSNMDCSMIALPVGGAHIOFLNFBTEBNHDIYELRNPYETSRMAGFSGS 1800  
 DB 1741 PGFPNGYPSNMDCSMIALPVGGAHIOFLNFBTEBNHDIYELRNPYETSRMAGFSGS 1800  
 QY 1801 ELPSLSLSTSHETTVYFHSDSQGNRPFKLEYQAYELOECBPDEPPANGIVRAGYVNGQ 1860  
 DB 1801 ELPSLSLSTSHETTVYFHSDSQGNRPFKLEYQAYELOECBPDEPPANGIVRAGYVNGQ 1860  
 QY 1861 SVTFECLPGYQLTGHVVLTCORGTNRNMDHPLPKCEVPGCGNITSSNGTYVSFSPSPVS 1920  
 DB 1861 SVTFECLPGYQLTGHVVLTCORGTNRNMDHPLPKCEVPGCGNITSSNGTYVSFSPSPVS 1920  
 QY 1921 SSODCVMLITVPIGHGVRLNLSLLOTEPSGDFITIDGPOQTAPRLGVFTRSMAKKTVO 1980  
 DB 1921 SSODCVMLITVPIGHGVRLNLSLLOTEPSGDFITIDGPOQTAPRLGVFTRSMAKKTVO 1980  
 QY 1981 SSNOVLLKPHRDAATGIFALIAFSAVPLTKCPPTLLPNAEVVTENEEFNIGDIIVRYCL 2040  
 DB 1981 SSNOVLLKPHRDAATGIFALIAFSAVPLTKCPPTLLPNAEVVTENEEFNIGDIIVRYCL 2040  
 QY 2041 PGFTLVGNELTKKCTGYLOFEGPBPICVHACTNBLLDSTVILISOSIPGYPQOTC 2100  
 DB 2041 PGFTLVGNELTKKCTGYLOFEGPBPICVHACTNBLLDSTVILISOSIPGYPQOTC 2100  
 QY 2101 SMLVRVBPYNISLTYEYFSEKQYDEFEIPDPSQOSPILKALSGNYSAPLIVTSSNS 2160  
 DB 2101 SMLVRVBPYNISLTYEYFSEKQYDEFEIPDPSQOSPILKALSGNYSAPLIVTSSNS 2160  
 QY 2161 VYLRWSSDHAYNKRGKIRYSAPYCSLPAPLHGFILGOSTPGGSIHFGNAGYRLVG 2220  
 DB 2161 VYLRWSSDHAYNKRGKIRYSAPYCSLPAPLHGFILGOSTPGGSIHFGNAGYRLVG 2220  
 QY 2221 HSMALICTRHPQGHLMSEAPILCOALSCEIPAPKXGMFGKTYVTGTAVYSCSGYHL 2280  
 DB 2221 HSMALICTRHPQGHLMSEAPILCOALSCEIPAPKXGMFGKTYVTGTAVYSCSGYHL 2280  
 QY 2281 OAGABATBECLDGTGLMNNRVPPQCVPTCCPDVSSISVEHGRRLIFETOYOPQAOQLMI 2340  
 DB 2281 OAGABATBECLDGTGLMNNRVPPQCVPTCCPDVSSISVEHGRRLIFETOYOPQAOQLMI 2340  
 QY 2341 CDPGYVYTGQVRVRCOANGKMSLGDSTPCRILISCEGLPIPNGBHIGTLVYGATAIFS 2400  
 DB 2341 CDPGYVYTGQVRVRCOANGKMSLGDSTPCRILISCEGLPIPNGBHIGTLVYGATAIFS 2400  
 QY 2401 CMSGYTLVGSRVRECMANGMSSEYRCLAGHGCTBEPIYNGHINBENTSYRSGVYQCN 2460  
 DB 2401 CMSGYTLVGSRVRECMANGMSSEYRCLAGHGCTBEPIYNGHINBENTSYRSGVYQCN 2460  
 QY 2461 AGFRLIGMSVRIQOQDHMSGKTPFCVPTICGHPGVNGLTOGNOFNLDVVKFCNPG 2520  
 DB 2461 AGFRLIGMSVRIQOQDHMSGKTPFCVPTICGHPGVNGLTOGNOFNLDVVKFCNPG 2520  
 QY 2521 YMAEGAARSGCLASGQMSDMLPTCRIINCTDPGHQENSYRQVHVASGPHRPSFGTIVSYRC 2580  
 DB 2521 YMAEGAARSGCLASGQMSDMLPTCRIINCTDPGHQENSYRQVHVASGPHRPSFGTIVSYRC 2580  
 QY 2581 NHGFYLLGTPTVLSQOQDGTDRRPOCLVSCGHBPSSPHSOMSGSYTYGAVVRYSCIG 2640  
 DB 2581 NHGFYLLGTPTVLSQOQDGTDRRPOCLVSCGHBPSSPHSOMSGSYTYGAVVRYSCIG 2640  
 QY 2641 KRTLVGNSTRMGLDGHMTGSLPHCSGTSVGVCGDPCGIPAHGIRLDSFDPGTVMFSC 2700

```

Db      2641 KRTLVGNSTRMCGJGDHWTGSLPHCSGTSVGVCCDPCIPIANGILGSPDPGTMRSC 2700
Qy      2701 AGHYLRGSSERTCOANGSGSGOPECCGYSCGNPGTBNARVFSDLVFSSTIVYECRE 2760
Db      2701 AGHYLRGSSERTCOANGSGSGOPECCGYSCGNPGTBNARVFSDLVFSSTIVYECRE 2760
Qy      2761 GYVATGLLSRHCSVNGTGTSPDCLVYNGCDPCIPIANGILGNDPFXNTVTMYQCPCGY 2820
Db      2761 GYVATGLLSRHCSVNGTGTSPDCLVYNGCDPCIPIANGILGNDPFXNTVTMYQCPCGY 2820
Qy      2821 MMEHRSVLSCTYDRRTWNGTKPVCKALMCKPPLIINGKVGSGDFMGGSSVTYACLEGY 2880
Db      2821 MMEHRSVLSCTYDRRTWNGTKPVCKALMCKPPLIINGKVGSGDFMGGSSVTYACLEGY 2880
Qy      2881 QLSLPAYFTCGNSGSMGELPQCPCPVCGDGVGVSRRGRDRGCSYSSVSFSCHPPLVL 2940
Db      2881 QLSLPAYFTCGNSGSMGELPQCPCPVCGDGVGVSRRGRDRGCSYSSVSFSCHPPLVL 2940
Qy      2941 VGSFRRFCQSDGTWGTGTOPSCIDPTLTTCADPGVPGVPGIIONNSQGVGSTVLFRCKGY 3000
Db      2941 VGSFRRFCQSDGTWGTGTOPSCIDPTLTTCADPGVPGVPGIIONNSQGVGSTVLFRCKGY 3000
Qy      3001 LLOSTRTCTLPNLTWGTGTPDDCVPHHCROPEPTTHANVAGALDLPMSGYTLITPARASP 3060
Db      3001 LLOSTRTCTLPNLTWGTGTPDDCVPHHCROPEPTTHANVAGALDLPMSGYTLITPARASP 3060
Qy      3061 SRVAPSTAPARMAAGASRPSAMRSGVDPSTPLPSHSPKP 3104
Db      3061 SRVAPSTAPARMAAGASRPSAMRSGVDPSTPLPSHSPKP 3104

RESULT 3
ADH71138
ID ADH71138 standard; protein; 3104 AA.
XX
XX ADH71138;
XX
XX 25-MAR-2004 (first entry)
XX
XX Human protein of the invention NOVab SEQ ID NO:34.
XX
XX human; cytostatic; immunomodulator; neuroprotective; nootropic;
XX anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;
XX vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
XX obesity; diabetes; infectious disease; metabolic syndrome X;
XX dyslipidaemia.
XX
XX Homo sapiens.
XX
XX MO2003102155-A2.
XX
XX 11-DEC-2003.
XX
XX 03-JUN-2003; 2003MO-US017430.
XX
XX 03-JUN-2002; 2002US-0385120P.
XX 04-JUN-2002; 2002US-0385784P.
XX 05-JUN-2002; 2002US-0386041P.
XX 06-JUN-2002; 2002US-0386476P.
XX 06-JUN-2002; 2002US-0386533P.
XX 06-JUN-2002; 2002US-0386643P.
XX 06-JUN-2002; 2002US-0387016P.
XX 07-JUN-2002; 2002US-0386796P.
XX 07-JUN-2002; 2002US-0386816P.
XX 07-JUN-2002; 2002US-0386931P.
XX 07-JUN-2002; 2002US-0386942P.
XX 07-JUN-2002; 2002US-0386971P.
XX 07-JUN-2002; 2002US-0387262P.
XX 08-JUN-2002; 2002US-0296960P.
XX 10-JUN-2002; 2002US-0387400P.
XX 10-JUN-2002; 2002US-0387535P.

```

```

PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387625P.
PR 11-JUN-2002; 2002US-0387634P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387702P.
PR 11-JUN-2002; 2002US-0387736P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387933P.
PR 12-JUN-2002; 2002US-0387946P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388056P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389118P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389144P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389729P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389848P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390209P.
PR 21-JUN-2002; 2002US-0390763P.
PR 17-JUL-2002; 2002US-0396706P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402156P.
PR 09-AUG-2002; 2002US-0402256P.
PR 12-AUG-2002; 2002US-0402389P.
PR 12-AUG-2002; 2002US-0402786P.
PR 12-AUG-2002; 2002US-0402816P.
PR 12-AUG-2002; 2002US-0402821P.
PR 13-AUG-2002; 2002US-0402832P.
PR 13-AUG-2002; 2002US-0403448P.
PR 13-AUG-2002; 2002US-0403459P.
PR 13-AUG-2002; 2002US-0403531P.
PR 13-AUG-2002; 2002US-0403532P.
PR 13-AUG-2002; 2002US-0403563P.
PR 15-AUG-2002; 2002US-0406317P.
PR 26-AUG-2002; 2002US-0403617P.
PR 26-AUG-2002; 2002US-0406182P.
PR 27-AUG-2002; 2002US-0406355P.
PR 27-AUG-2002; 2002US-0406240P.
PR 12-SEP-2002; 2002US-0410084P.
PR 20-SEP-2002; 2002US-0412528P.
PR 23-SEP-2002; 2002US-0412731P.
PR 30-SEP-2002; 2002US-0414801P.
PR 30-SEP-2002; 2002US-0414839P.
PR 30-SEP-2002; 2002US-0414840P.
PR 30-SEP-2002; 2002US-0414954P.
PR 09-OCT-2002; 2002US-0417186P.
PR 09-OCT-2002; 2002US-0417406P.
PR 23-OCT-2002; 2002US-0420639P.
PR 28-OCT-2002; 2002US-0421156P.
PR 31-OCT-2002; 2002US-0422690P.
PR 01-NOV-2002; 2002US-0423130P.
PR 05-NOV-2002; 2002US-0423798P.
PR 12-NOV-2002; 2002US-0425453P.

(CURA-) CURAGEN CORP.
XX
XX Alsebrook JP, Alvarez E, Anderson DW, Boldog FL, Casman SJ;
XX Catteiron E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
XX Ettenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
XX Guev YI, Herrmann UL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR;
XX MacIsachlan T, Malpankar UM, Mezick AJ, Millet I, Mishra VS;
XX Padigara M, Patturajan M, Pena CBA, Peyman JA, Raha D, Rastelli L;
XX Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkets RA;
XX Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;
XX Zhong H;
XX
XX WPI; 2004-081935/08.

```

DR N-PSDB; ADH71137.  
 XX New NOVX polypeptides and nucleic acid molecules useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or  
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.  
 XX  
 XX Example 4; SEQ ID NO 34; 1880bp; English.  
 CC The invention relates to a novel isolated polypeptide (NOVX). A  
 CC polypeptide of the invention has cytostatic, immunomodulator,  
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and  
 CC antihypertensive activity, and may have a use in gene therapy, and as a  
 CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising  
 CC any of the 303 fully defined nucleotide sequences given in the  
 CC specification. The polypeptide is useful in the manufacture of a  
 CC medicament for treating a syndrome associated with a human disease. The  
 CC polypeptide, polynucleotide and antibody are useful in diagnosing,  
 CC treating or preventing NOVX-associated disorders, e.g. cancer, cachexia,  
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious  
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are  
 CC further used as hybridisation probes, in chromosome mapping, tissue  
 CC typing, preventive medicine, and pharmacogenomics. The present sequence  
 CC represents a NOVX polypeptide of the invention.  
 CC  
 XX  
 XX Sequence 3104 AA:  
 Query Match 100.0%; Score 16385; DB 8; Length 3104;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 661 GEPLECEPNOWSRALPSCALCGFLOQSSGTLSPGPPDPYNNUNCTMIETSHGK 720  
 DB 661 GEPLECEPNOWSRALPSCALCGFLOQSSGTLSPGPPDPYNNUNCTMIETSHGK 720  
 QY 721 VPFTHFTLHESGDYLLITENGSTPQPLRLTGSRLPAPISAGLYGNFTAOVRFLSDPS 780  
 DB 721 VPFTHFTLHESGDYLLITENGSTPQPLRLTGSRLPAPISAGLYGNFTAOVRFLSDPS 780  
 QY 781 MSYEGNITTSBYDLEPCBPBPAYSIKGLQFGYDITLTCSCFPGYRLGRTARITCLG 840  
 DB 781 MSYEGNITTSBYDLEPCBPBPAYSIKGLQFGYDITLTCSCFPGYRLGRTARITCLG 840  
 QY 841 GRRRLMSSPLPRCAVACGNSVTGQTLSPNPNVNNHETIYSIQTPGKGIOLKAR 900  
 DB 841 GRRRLMSSPLPRCAVACGNSVTGQTLSPNPNVNNHETIYSIQTPGKGIOLKAR 900  
 QY 901 AFELSGDYLVKYDGNNSARLLGVFSHSEMGVTINSSSLMLDPTDAENTSKGFEL 960  
 DB 901 AFELSGDYLVKYDGNNSARLLGVFSHSEMGVTINSSSLMLDPTDAENTSKGFEL 960  
 QY 961 HFSFELIKCEDPTKPGYKYHDEBGFAGSSVSFSCDPGYSIRGSEBELCLSGERTWD 1020  
 DB 961 HFSFELIKCEDPTKPGYKYHDEBGFAGSSVSFSCDPGYSIRGSEBELCLSGERTWD 1020  
 QY 1021 RPLPTCAVECGGTVRBEVSGVLTSPGYPAPEHNLNLCIMTEACCTGLHFLVPTDEE 1080  
 DB 1021 RPLPTCAVECGGTVRBEVSGVLTSPGYPAPEHNLNLCIMTEACCTGLHFLVPTDEE 1080  
 QY 1081 VHDVLRIMDGPVSSGVLLKELSGPALPKDLHSTFNSVVLQFSTDFPFSKGFALIOFSVST 1140  
 DB 1081 VHDVLRIMDGPVSSGVLLKELSGPALPKDLHSTFNSVVLQFSTDFPFSKGFALIOFSVST 1140  
 QY 1141 ATSCNDPGLPONGSRSGDSHAGDSTVPQCDPGYALQGSABEISCVKIEENRFTMQPSPTC 1200  
 DB 1141 ATSCNDPGLPONGSRSGDSHAGDSTVPQCDPGYALQGSABEISCVKIEENRFTMQPSPTC 1200  
 QY 1201 IAPCGDILGPGSVLTSPNYPPEYPPGKEDMKVTVSPDVIALVNIINFLBEGYFLHI 1260  
 DB 1201 IAPCGDILGPGSVLTSPNYPPEYPPGKEDMKVTVSPDVIALVNIINFLBEGYFLHI 1260  
 QY 1261 YDGRDLSPLIGSFYSGQLPGRTESSNSLFLAFRSDASVSNAGFYIDYENRRESCFDP 1320  
 DB 1261 YDGRDLSPLIGSFYSGQLPGRTESSNSLFLAFRSDASVSNAGFYIDYENRRESCFDP 1320  
 QY 1321 GSINKGTRVGSDLKLGSSVTTYCHGSEYVETSTLSCILGPDGKPVWNNRPLVCTAPCGG 1380  
 DB 1321 GSINKGTRVGSDLKLGSSVTTYCHGSEYVETSTLSCILGPDGKPVWNNRPLVCTAPCGG 1380  
 QY 1381 QYVSGDVLTSPNYPNTSGOICLYFVTVPKDYVYFGQAFPHTLANDVVEVHDGSHQ 1440  
 DB 1381 QYVSGDVLTSPNYPNTSGOICLYFVTVPKDYVYFGQAFPHTLANDVVEVHDGSHQ 1440  
 QY 1441 SRLSSLSGSHGTESEPLATSNQVLLKFSAGKGLAPRGHFVYQAVPTSATQCSSVPER 1500  
 DB 1441 SRLSSLSGSHGTESEPLATSNQVLLKFSAGKGLAPRGHFVYQAVPTSATQCSSVPER 1500  
 QY 1501 RYKRLGSDPFSVGAIVRFECNSGYALQGSBEIECLVPALAQMNVSAPTCVPPCGANLT 1560  
 DB 1501 RYKRLGSDPFSVGAIVRFECNSGYALQGSBEIECLVPALAQMNVSAPTCVPPCGANLT 1560  
 QY 1561 ERRGTLISGFPPEPYNSLNCWKIIVPBGAGIQIOVVSFVTRQNNDSLEVFQGANVTY 1620  
 DB 1561 ERRGTLISGFPPEPYNSLNCWKIIVPBGAGIQIOVVSFVTRQNNDSLEVFQGANVTY 1620  
 QY 1621 MLGSPSGGTVPALNLTNSQVLLHFYSDSISVSAAGFHELYKTVGLSSCEBPAPPSNGVXT 1680  
 DB 1621 MLGSPSGGTVPALNLTNSQVLLHFYSDSISVSAAGFHELYKTVGLSSCEBPAPPSNGVXT 1680  
 QY 1681 GERLYLVNDVVSFOCEBPGYALQGHAIISCMPTGYRRNNYPPPLCIAOCGGTVEBMEGVILS 1740  
 DB 1681 GERLYLVNDVVSFOCEBPGYALQGHAIISCMPTGYRRNNYPPPLCIAOCGGTVEBMEGVILS 1740

```

QY 1741 PGFGPNYPSNMDCKWIALPYVFGAHIQFLNFSTEPNHDYIEIRNGPEYTSRMKGRSGS 1800
DB 1741 PGFGPNYPSNMDCKWIALPYVFGAHIQFLNFSTEPNHDYIEIRNGPEYTSRMKGRSGS 1800
QY 1801 ELPSLSLSTSHETTYYFPHSDHSQNRPGFKLEYAYELOECDEPPEFANGIVRGAGYVWQ 1860
DB 1801 ELPSLSLSTSHETTYYFPHSDHSQNRPGFKLEYAYELOECDEPPEFANGIVRGAGYVWQ 1860
QY 1861 SYTEECIPGYOLTGHPVLTQCHGTRNMDHDLPCCEVPCCGNITSSNGITYSFGPSPPYS 1920
DB 1861 SYTEECIPGYOLTGHPVLTQCHGTRNMDHDLPCCEVPCCGNITSSNGITYSFGPSPPYS 1920
QY 1921 SSODCWMLITVPIGHVRLNLSLQTEPSCGFITIMDGPQOTAPRLGVEFTRMAKKTIVOS 1980
DB 1921 SSODCWMLITVPIGHVRLNLSLQTEPSCGFITIMDGPQOTAPRLGVEFTRMAKKTIVOS 1980
QY 1981 SSNOVLLKFRHDAATGGI PAIAFSAVPLTKCPPTIIPNAEVTENEENIGDIVRYCRL 2040
DB 1981 SSNOVLLKFRHDAATGGI PAIAFSAVPLTKCPPTIIPNAEVTENEENIGDIVRYCRL 2040
QY 2041 PGFTLVGNELITCKLGTLYLOFEGBPPICEVHCPTNELITDSTGVILISQSYPSGYPOFC 2100
DB 2041 PGFTLVGNELITCKLGTLYLOFEGBPPICEVHCPTNELITDSTGVILISQSYPSGYPOFC 2100
QY 2101 SWLVREPDYNI SLTVEYFLSEKQYDFEIDFGPSGSPILKALSGNYSAPLIVTSSNS 2160
DB 2101 SWLVREPDYNI SLTVEYFLSEKQYDFEIDFGPSGSPILKALSGNYSAPLIVTSSNS 2160
QY 2161 VYLWSSDHAIVNKGFEIRYSAPYCSLPAPLHGFILIGQSTPGGSIHFGCNAGYRLVG 2220
DB 2161 VYLWSSDHAIVNKGFEIRYSAPYCSLPAPLHGFILIGQSTPGGSIHFGCNAGYRLVG 2220
QY 2221 HSNALCTRHPOGYHLMSEALPLCOALSCGLPEAPKGMVGEKYTGKAVYSCSEGYH 2280
DB 2221 HSNALCTRHPOGYHLMSEALPLCOALSCGLPEAPKGMVGEKYTGKAVYSCSEGYH 2280
QY 2281 QAQAERTAECLDTGLMSNRNVPQCVPTCPDVSSISVEHGRMLIFETQYQFOAQLMI 2340
DB 2281 QAQAERTAECLDTGLMSNRNVPQCVPTCPDVSSISVEHGRMLIFETQYQFOAQLMI 2340
QY 2341 CDPGYVYTGORVIRCOANGKMSLGDSTPTCRITISCELPPIPNHRICTLSTVYGAAIATS 2400
DB 2341 CDPGYVYTGORVIRCOANGKMSLGDSTPTCRITISCELPPIPNHRICTLSTVYGAAIATS 2400
QY 2401 CNGSYTLVGSRYVRECMANGLMSGSEVRCLAGHGTPEPIVNGHINENSYSGSVYQCN 2460
DB 2401 CNGSYTLVGSRYVRECMANGLMSGSEVRCLAGHGTPEPIVNGHINENSYSGSVYQCN 2460
QY 2461 AGFRLIGMSVRIQOQDHHSWGKTPFCVPTTCGHPGNPVNGLTOGNOFNINDVVKFYCNG 2520
DB 2461 AGFRLIGMSVRIQOQDHHSWGKTPFCVPTTCGHPGNPVNGLTOGNOFNINDVVKFYCNG 2520
QY 2521 YMAEGARSOCLASGQMDMLPTCRIINCTDPGHQNSROYHASGPHRFSGTTVSYNC 2580
DB 2521 YMAEGARSOCLASGQMDMLPTCRIINCTDPGHQNSROYHASGPHRFSGTTVSYNC 2580
QY 2581 NHGFYLLGTPVLSQOQDGTWDRPPOCLIVSCGHPSPHSQMSGDSYTVGAVVYXSCIG 2640
DB 2581 NHGFYLLGTPVLSQOQDGTWDRPPOCLIVSCGHPSPHSQMSGDSYTVGAVVYXSCIG 2640
QY 2641 KRTLVGNSTRMGLDGHMTGSLPHGSGTVGCGDGTIAHGRIRLGDSDPCTVWRFSCE 2700
DB 2641 KRTLVGNSTRMGLDGHMTGSLPHGSGTVGCGDGTIAHGRIRLGDSDPCTVWRFSCE 2700
QY 2701 AGHVLAGSSERTCOANGSWSGQPECGVISCNPGTSPNARVAFSDGLVSSSIYECRE 2760
DB 2701 AGHVLAGSSERTCOANGSWSGQPECGVISCNPGTSPNARVAFSDGLVSSSIYECRE 2760
QY 2761 GYATATLLLRHGSVNGTWTGSDPECLVINGCPGIPANGRLIGDPRYKTKVTVYQCVGY 2820
DB 2761 GYATATLLLRHGSVNGTWTGSDPECLVINGCPGIPANGRLIGDPRYKTKVTVYQCVGY 2820
QY 2821 MMESHRSVSLSCRTKRYNKGTPVCKALMCKRPPILPNGKVVSGSPMMGSSVTVACLEGY 2880

```

```

DB 2821 MMESHRSVSLSCRTKRYNKGTPVCKALMCKRPPILPNGKVVSGSPMMGSSVTVACLEGY 2880
QY 2881 QLSLPAVFTCEANGSWTGLPQCFPVFCGDPVPSRGREDGFSYRSVSSCHPPLVL 2940
DB 2881 QLSLPAVFTCEANGSWTGLPQCFPVFCGDPVPSRGREDGFSYRSVSSCHPPLVL 2940
QY 2941 VSPRRFCOSDGTWSGTQPSCIDPILTTCCADBGVQFGIQQNSQGYQVGSYVLPFCQKGY 3000
DB 2941 VSPRRFCOSDGTWSGTQPSCIDPILTTCCADBGVQFGIQQNSQGYQVGSYVLPFCQKGY 3000
QY 3001 LLOGSTTRTCLPNLFWSGPPDCVPHHCRQPTPHAVGALDLPMSGYTLITPARASP 3060
DB 3001 LLOGSTTRTCLPNLFWSGPPDCVPHHCRQPTPHAVGALDLPMSGYTLITPARASP 3060
QY 3061 SRVASTAPARRMAGQASRPSAMRSRGVGDPSLPGSHRSBKP 3104
DB 3061 SRVASTAPARRMAGQASRPSAMRSRGVGDPSLPGSHRSBKP 3104

RESULT 4
ADH71168
ID ADH71168 standard; protein; 3104 AA.
XX
AC ADH71168;
XX
DT 25-MAR-2004 (first entry)
XX
DE Human protein of the invention NOV4q SEQ ID NO:64.
XX
KW human; cytosolic; immunomodulator; neuroprotective; nootropic;
KW anorectic; antidiabetic; antimicrobial; antilipaeitic; gene therapy;
KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
KW obesity; diabetes; infectious disease; metabolic syndrome X;
KW dyslipidaemia.
XX
OS Homo sapiens.
XX
PN W02003102155-A2.
XX
PD 11-DEC-2003.
XX
PF 03-JUN-2003; 2003WC-US017430.
XX
PR 03-JUN-2002; 2002US-0385120P.
PR 04-JUN-2002; 2002US-0385784P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386047P.
PR 06-JUN-2002; 2002US-0386453P.
PR 06-JUN-2002; 2002US-0386864P.
PR 06-JUN-2002; 2002US-0387016P.
PR 07-JUN-2002; 2002US-038796P.
PR 07-JUN-2002; 2002US-038816P.
PR 07-JUN-2002; 2002US-0389312P.
PR 07-JUN-2002; 2002US-038942P.
PR 07-JUN-2002; 2002US-0389711P.
PR 07-JUN-2002; 2002US-0387262P.
PR 08-JUN-2002; 2002US-028696P.
PR 10-JUN-2002; 2002US-0387400P.
PR 10-JUN-2002; 2002US-0387555P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387625P.
PR 11-JUN-2002; 2002US-0387634P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387686P.
PR 11-JUN-2002; 2002US-0387702P.
PR 11-JUN-2002; 2002US-0387836P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387933P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.

```

PR 12-JUN-2002; 2002US-0388036P.  
 PR 13-JUN-2002; 2002US-0389123P.  
 PR 14-JUN-2002; 2002US-0389118P.  
 PR 14-JUN-2002; 2002US-0389120P.  
 PR 14-JUN-2002; 2002US-0389146P.  
 PR 14-JUN-2002; 2002US-0389146P.  
 PR 17-JUN-2002; 2002US-0389729P.  
 PR 17-JUN-2002; 2002US-0389742P.  
 PR 18-JUN-2002; 2002US-0389884P.  
 PR 19-JUN-2002; 2002US-0390006P.  
 PR 19-JUN-2002; 2002US-0390209P.  
 PR 21-JUN-2002; 2002US-0390763P.  
 PR 17-JUL-2002; 2002US-0396706P.  
 PR 06-AUG-2002; 2002US-0401628P.  
 PR 09-AUG-2002; 2002US-0402156P.  
 PR 09-AUG-2002; 2002US-040256P.  
 PR 09-AUG-2002; 2002US-0402389P.  
 PR 12-AUG-2002; 2002US-0402786P.  
 PR 12-AUG-2002; 2002US-0402816P.  
 PR 12-AUG-2002; 2002US-0402821P.  
 PR 12-AUG-2002; 2002US-0402832P.  
 PR 13-AUG-2002; 2002US-0403448P.  
 PR 13-AUG-2002; 2002US-0403459P.  
 PR 13-AUG-2002; 2002US-0403531P.  
 PR 13-AUG-2002; 2002US-0403532P.  
 PR 13-AUG-2002; 2002US-0403563P.  
 PR 13-AUG-2002; 2002US-0406317P.  
 PR 15-AUG-2002; 2002US-0403617P.  
 PR 26-AUG-2002; 2002US-0406182P.  
 PR 26-AUG-2002; 2002US-0406355P.  
 PR 27-AUG-2002; 2002US-0406240P.  
 PR 12-SEP-2002; 2002US-0410084P.  
 PR 20-SEP-2002; 2002US-0412528P.  
 PR 23-SEP-2002; 2002US-0412731P.  
 PR 30-SEP-2002; 2002US-0414801P.  
 PR 30-SEP-2002; 2002US-0414839P.  
 PR 30-SEP-2002; 2002US-0414840P.  
 PR 30-SEP-2002; 2002US-0414954P.  
 PR 09-OCT-2002; 2002US-0417186P.  
 PR 09-OCT-2002; 2002US-0417406P.  
 PR 23-OCT-2002; 2002US-0420639P.  
 PR 28-OCT-2002; 2002US-0421156P.  
 PR 31-OCT-2002; 2002US-0422690P.  
 PR 01-NOV-2002; 2002US-0423130P.  
 PR 05-NOV-2002; 2002US-0423798P.  
 PR 05-NOV-2002; 2002US-0423798P.  
 PR 12-NOV-2002; 2002US-0425453P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Alabrook JP, Alvarez E, Anderson DW, Boldog FL, Caeman SJ;  
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;  
 PI Eberstein S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;  
 PI Gusev VY, Hermann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR;  
 PI MacLachlan T, Malvanter UM, Mezick AJ, Millet I, Mishra VS;  
 PI MacLachlan T, Malvanter UM, Mezick AJ, Millet I, Mishra VS;  
 PI Rieger DK, Rotenberg ME, Scioe P, Shenoy SG, Shimkets RA;  
 PI Smithson G, Spletter KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;  
 PI Zhong H;  
 XX  
 DR WPI: 2004-081935/08.  
 DR N-PSDB; ADH71167.  
 XX  
 PT New NOVX polypeptides and nucleic acid molecules useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or  
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.  
 XX  
 SS Example 4, SEQ ID NO 64; 1880bp; English.  
 XX  
 CC The invention relates to a novel isolated polypeptide (NOVX). A  
 CC polypeptide of the invention has cytostatic, immunomodulator,  
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and  
 CC antilipemic activity, and may have a use in gene therapy, and as a

CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising  
 CC any of the 303 fully defined nucleotide sequences given in the  
 CC specification. The polypeptide is useful in the manufacture of a  
 CC medicament for treating a syndrome associated with a human disease. The  
 CC polypeptide, polynucleotide and antibody are useful in diagnosing,  
 CC treating or preventing NOVX-associated disorders, e.g. cancer, cachexia,  
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious  
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are  
 CC further used as hybridisation probes, in chromosome mapping, tissue  
 CC typing, preventive medicine, and pharmacogenomics. The present sequence  
 CC represents a NOVX polypeptide of the invention.  
 XX  
 SQ Sequence 3104 AA;  
 Query Match 100.0%; Score 16978; DB 8; Length 3104;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3103; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MAGAPPALLPCLSLSDCCASNORHVGVPSELVKKQJELKSRGVKLMPSKDNISQKTS 60  
 DB 1 MAGAPPALLPCLSLSDCCASNORHVGVPSELVKKQJELKSRGVKLMPSKDNISQKTS 60  
 QY VLTQVGSQGHNNCPDGPGERGRGLSDRLGSSVQFTCNBEGYDLOGSKRITCMKVS DM 120  
 DB VLTQVGSQGHNNCPDGPGERGRGLSDRLGSSVQFTCNBEGYDLOGSKRITCMKVS DM 120  
 QY 61 VLTQVGSQGHNNCPDGPGERGRGLSDRLGSSVQFTCNBEGYDLOGSKRITCMKVS DM 120  
 DB 61 VLTQVGSQGHNNCPDGPGERGRGLSDRLGSSVQFTCNBEGYDLOGSKRITCMKVS DM 120  
 QY 121 FAAMSDHPRVCRARMCDAHRLRSGGIIITSNFPIQDNNHACWIIITALLMPSKYIKLAF 180  
 DB 121 FAAMSDHPRVCRARMCDAHRLRSGGIIITSNFPIQDNNHACWIIITALLMPSKYIKLAF 180  
 QY 181 EFDLERYDTLVTDGQGDQDQKTVLMSQMAQSDSHHTGSRIPESMSGDIIMROKTVL 240  
 DB 181 EFDLERYDTLVTDGQGDQDQKTVLMSQMAQSDSHHTGSRIPESMSGDIIMROKTVL 240  
 QY 181 EFDLERYDTLVTDGQGDQDQKTVLMSQMAQSDSHHTGSRIPESMSGDIIMROKTVL 240  
 DB 181 EFDLERYDTLVTDGQGDQDQKTVLMSQMAQSDSHHTGSRIPESMSGDIIMROKTVL 240  
 QY 241 EICRDISSDARSQSVKSPKTSNAVELVAPGTEIEGSGCDPGIIPAYGRREGSRFHGD 300  
 DB 241 EICRDISSDARSQSVKSPKTSNAVELVAPGTEIEGSGCDPGIIPAYGRREGSRFHGD 300  
 QY 301 TLKFECPAPFELVGOKAITQKNNQSAKAPGVCSFFNFPTSPSGVVLSPNYPEDYGNH 360  
 DB 301 TLKFECPAPFELVGOKAITQKNNQSAKAPGVCSFFNFPTSPSGVVLSPNYPEDYGNH 360  
 QY 361 LHCWMLILARPEERIHAFNDIVPEQDFLVKDGATAAPVLGTPSGNQLPSSITSSG 420  
 DB 361 LHCWMLILARPEERIHAFNDIVPEQDFLVKDGATAAPVLGTPSGNQLPSSITSSG 420  
 QY 421 HVARLEFQDHTSGKRGFNITFTFRNECPDGPVPVNGKRFQDSIQLGSSISFLCDEGF 480  
 DB 421 HVARLEFQDHTSGKRGFNITFTFRNECPDGPVPVNGKRFQDSIQLGSSISFLCDEGF 480  
 QY 481 LGTQGSSTITCVLKESGVVNSAVLCEAPCGHLISPSGIIISPGMPGYKDALSCAMV 540  
 DB 481 LGTQGSSTITCVLKESGVVNSAVLCEAPCGHLISPSGIIISPGMPGYKDALSCAMV 540  
 QY 541 IBAQPGYKITDRPFTENYVDTLEVRDRTYSAPIIGVYHGVQVPLISNLYL 600  
 DB 541 IBAQPGYKITDRPFTENYVDTLEVRDRTYSAPIIGVYHGVQVPLISNLYL 600  
 QY 601 FSTDKSHSDIGPOLRYETITLQSDHCLDPGIPVNGRQHGNDFFVGLVTFSCSGYTLSD 660  
 DB 601 FSTDKSHSDIGPOLRYETITLQSDHCLDPGIPVNGRQHGNDFFVGLVTFSCSGYTLSD 660  
 QY 661 GEPLCEPNQWMBALPSCGALGGGFGSSGIIISGFPDFPNNINCTWIIITSHGKG 720  
 DB 661 GEPLCEPNQWMBALPSCGALGGGFGSSGIIISGFPDFPNNINCTWIIITSHGKG 720  
 QY 721 VPFTHFHEHESGDYLLITENGSPLOPLQGLSRIPAPISAGLGNPFAQVRFISDFS 780  
 DB 721 VPFTHFHEHESGDYLLITENGSPLOPLQGLSRIPAPISAGLGNPFAQVRFISDFS 780  
 QY 781 MSYEGFNITSEYDLCEBCEPEVPAYSIRKGLQGVGDTLVFSCFPGRYLEGARITCLG 840  
 DB 781 MSYEGFNITSEYDLCEBCEPEVPAYSIRKGLQGVGDTLVFSCFPGRYLEGARITCLG 840

QY 841 GRRRLMSPLPRCVAECGNSVTGOTLLSPNFPVNNNNHCTYSITOTPOCKIOLKAR 900  
DB 841 GRRRLMSPLPRCVAECGNSVTGOTLLSPNFPVNNNNHCTYSITOTPOCKIOLKAR 900  
QY 901 AFELSEGDLKYYDGNNSARLLGVFSEMMGVTLNSTSSLMDFITDAENTSKGPEL 960  
DB 901 AFELSEGDLKYYDGNNSARLLGVFSEMMGVTLNSTSSLMDFITDAENTSKGPEL 960  
QY 961 HPSSELLKCEBPGPKFGYVHDBGHFAGSSVSCDPGSLNGSEELLCLSGERTWD 1020  
DB 961 HPSSELLKCEBPGPKFGYVHDBGHFAGSSVSCDPGSLNGSEELLCLSGERTWD 1020  
QY 1021 RPLPTCVAECGTVAGEVSGQVLSGYPAPYEHNLNCITWTEAAGCTIGLHFLVPTEE 1080  
DB 1021 RPLPTCVAECGTVAGEVSGQVLSGYPAPYEHNLNCITWTEAAGCTIGLHFLVPTEE 1080  
QY 1081 VHDVLRITWDGVEGVLKELISGPALPKDLHSTNSVVLQFSTDFTSKOGFAIQFSVST 1140  
DB 1081 VHDVLRITWDGVEGVLKELISGPALPKDLHSTNSVVLQFSTDFTSKOGFAIQFSVST 1140  
QY 1141 ATSCNDPGIPONGSRSGDSWAGSTVFCODPGYALOGSAEISGVKLENFPMQPSPTC 1200  
DB 1141 ATSCNDPGIPONGSRSGDSWAGSTVFCODPGYALOGSAEISGVKLENFPMQPSPTC 1200  
QY 1201 IAPCGDLTGSPSVILSPNYPEPYPKGECMDKVTSPDYIALVFNIFMLPEGYDPLH 1260  
DB 1201 IAPCGDLTGSPSVILSPNYPEPYPKGECMDKVTSPDYIALVFNIFMLPEGYDPLH 1260  
QY 1261 YDGRDLSPLIGSYGQOLPGRITSSNSLFLAFRSASVSNAGFYDYTENRESCFDP 1320  
DB 1261 YDGRDLSPLIGSYGQOLPGRITSSNSLFLAFRSASVSNAGFYDYTENRESCFDP 1320  
QY 1321 GSINGRVGSDDLKGSSTVYCHGGEVEGTSTLSCITLGPDPGPNVNNRPVCTAPCG 1380  
DB 1321 GSINGRVGSDDLKGSSTVYCHGGEVEGTSTLSCITLGPDPGPNVNNRPVCTAPCG 1380  
QY 1381 QYVGSDBVUSPNYPONTSGOICLYVTYPKDYVFGPAPFHTALNDVEVHDHSHOH 1440  
DB 1381 QYVGSDBVUSPNYPONTSGOICLYVTYPKDYVFGPAPFHTALNDVEVHDHSHOH 1440  
QY 1441 SRLLSLSGSHTEBSJPLATSNOLYKFSAKGLAPARGHRYVQVNPRTSATOCSSVPR 1500  
DB 1441 SRLLSLSGSHTEBSJPLATSNOLYKFSAKGLAPARGHRYVQVNPRTSATOCSSVPR 1500  
QY 1501 RYKRLGSDFSVGAIVAFECNSGYALQGSPEIECLVPGALQOMVSAFTCVPCGGLT 1560  
DB 1501 RYKRLGSDFSVGAIVAFECNSGYALQGSPEIECLVPGALQOMVSAFTCVPCGGLT 1560  
QY 1561 ERKGTLLSPGFPPEPYNLSNCWKIIVPBAGIIOIVVSVTEQOMDLSLEVFDGADNTVT 1620  
DB 1561 ERKGTLLSPGFPPEPYNLSNCWKIIVPBAGIIOIVVSVTEQOMDLSLEVFDGADNTVT 1620  
QY 1621 MGSFSGTTPALLNSTSNOLYHAFSDISVSAAGHLYEKTVGLSCDEBPVPSNGYKT 1680  
DB 1621 MGSFSGTTPALLNSTSNOLYHAFSDISVSAAGHLYEKTVGLSCDEBPVPSNGYKT 1680  
QY 1681 GERLYLVNDVVSFOCEBGYALQGHAIISCMPTVRANVYPPPLCIAQCGGTVEEMGVLLS 1740  
DB 1681 GERLYLVNDVVSFOCEBGYALQGHAIISCMPTVRANVYPPPLCIAQCGGTVEEMGVLLS 1740  
QY 1741 PGFPGYPSNMDCKMIALPVGGAHIOFLNSTENHNYIETRNGPYTSSMMGRFSGS 1800  
DB 1741 PGFPGYPSNMDCKMIALPVGGAHIOFLNSTENHNYIETRNGPYTSSMMGRFSGS 1800  
QY 1801 ELPPSLISTSHETTVFHSDBSONRPGFLEYQAEILOECPDEPFANGIYAGAGYNGO 1860  
DB 1801 ELPPSLISTSHETTVFHSDBSONRPGFLEYQAEILOECPDEPFANGIYAGAGYNGO 1860  
QY 1861 SVTFECLPBYOLTGHPVLTCHGTRNMDHPLPKCEVPCCGNITSSNGTVSPGPPSY 1920  
DB 1861 SVTFECLPBYOLTGHPVLTCHGTRNMDHPLPKCEVPCCGNITSSNGTVSPGPPSY 1920

QY 1921 SSODCWLLITVPIGHVRLNLSLQTEPSGDEITITWDGPOQAPRLGVFTRSMAKTVOS 1980  
DB 1921 SSODCWLLITVPIGHVRLNLSLQTEPSGDEITITWDGPOQAPRLGVFTRSMAKTVOS 1980  
QY 1981 SSNOVLKPHRAAAGCIFAIFSAVPLTKCPPTLTPAAEVTENEEFNIGDIYRYRCL 2040  
DB 1981 SSNOVLKPHRAAAGCIFAIFSAVPLTKCPPTLTPAAEVTENEEFNIGDIYRYRCL 2040  
QY 2041 PGFTLVGNELITCKLTGYLOFEGPPEICVHCPTEMLLTDSTGYLLSOSPGSYPOFC 2100  
DB 2041 PGFTLVGNELITCKLTGYLOFEGPPEICVHCPTEMLLTDSTGYLLSOSPGSYPOFC 2100  
QY 2101 SWLVREBPYNSLVEYFLSEKQYDEBEIFDGPSSGPLKALSGNSAPLIVTSSNS 2160  
DB 2101 SWLVREBPYNSLVEYFLSEKQYDEBEIFDGPSSGPLKALSGNSAPLIVTSSNS 2160  
QY 2161 VYLRMSDHAHYRKFXTIRYSAPYCSLRAPLHGTLTQTSPOGSGIHPGNACTYLVG 2220  
DB 2161 VYLRMSDHAHYRKFXTIRYSAPYCSLRAPLHGTLTQTSPOGSGIHPGNACTYLVG 2220  
QY 2221 HSMALCTHRPOGYHLSBAIPLCOALSCGLBPARGNVFGEKTYVTKAVYSCSEGYH 2280  
DB 2221 HSMALCTHRPOGYHLSBAIPLCOALSCGLBPARGNVFGEKTYVTKAVYSCSEGYH 2280  
QY 2281 QAGABATBCLDTGMSNRNVPPOCVPTGCDVSIISYEHGRMLIFETQYQFOALMLI 2340  
DB 2281 QAGABATBCLDTGMSNRNVPPOCVPTGCDVSIISYEHGRMLIFETQYQFOALMLI 2340  
QY 2341 CDPGYTYTGORVIRCOANGKSLGDSPTFCAIISCGELPIPNNGHRICTLTVYGATAIFS 2400  
DB 2341 CDPGYTYTGORVIRCOANGKSLGDSPTFCAIISCGELPIPNNGHRICTLTVYGATAIFS 2400  
QY 2401 CNSGTYVGSVREBMANGLSGSEBRTCLACHCTPEPIVNGHINGENYSYRSGVYQCN 2460  
DB 2401 CNSGTYVGSVREBMANGLSGSEBRTCLACHCTPEPIVNGHINGENYSYRSGVYQCN 2460  
QY 2461 AGFRLIGMSVRIICODHHMSGKTPFCVPTGHPGNPNGLTOGNOMTLNDVNFVCPNG 2520  
DB 2461 AGFRLIGMSVRIICODHHMSGKTPFCVPTGHPGNPNGLTOGNOMTLNDVNFVCPNG 2520  
QY 2521 YMAEGAARSQCLASGQWSDMLPTCRIINCTDPGHQENSVRQVHAGSHRFSFGTVSYRC 2580  
DB 2521 YMAEGAARSQCLASGQWSDMLPTCRIINCTDPGHQENSVRQVHAGSHRFSFGTVSYRC 2580  
QY 2581 NHGFYLLGTPLYSCOGGCTMDRPRPOCLLYSCGHPGSPHSONSGDSTYTGAVVRYCIG 2640  
DB 2581 NHGFYLLGTPLYSCOGGCTMDRPRPOCLLYSCGHPGSPHSONSGDSTYTGAVVRYCIG 2640  
QY 2641 KRTLVGNSTRMCGLDGHWMTSLPHCSGTSVGCDDPGI PAHGIRLIGDSFPGTYMRFSCE 2700  
DB 2641 KRTLVGNSTRMCGLDGHWMTSLPHCSGTSVGCDDPGI PAHGIRLIGDSFPGTYMRFSCE 2700  
QY 2701 AGHVLBSSSEPTCOANGSNMGSOPEGVISCNPGTFSNARVYFSDGLVFSSTIYBCE 2760  
DB 2701 AGHVLBSSSEPTCOANGSNMGSOPEGVISCNPGTFSNARVYFSDGLVFSSTIYBCE 2760  
QY 2761 GYATGLSHRCSYNGMTGSDPECLVINGDGIIPANGRLANDPFRYNTVYQCCPYG 2820  
DB 2761 GYATGLSHRCSYNGMTGSDPECLVINGDGIIPANGRLANDPFRYNTVYQCCPYG 2820  
QY 2821 MESHHRVSVLSTKDRMTNGTKVCAKLMCKPPLIPNGVYVSDTMMGSSVYALCLEG 2880  
DB 2821 MESHHRVSVLSTKDRMTNGTKVCAKLMCKPPLIPNGVYVSDTMMGSSVYALCLEG 2880  
QY 2881 QLSLPAFTCEBGNSWTGELPOCFPVPCDPPGVSGRREDRGFYSRSSVSFSCHPPLV 2940  
DB 2881 QLSLPAFTCEBGNSWTGELPOCFPVPCDPPGVSGRREDRGFYSRSSVSFSCHPPLV 2940  
QY 2941 VGSPPRFCCSDGTWSTGTOPSCIDPTLTTCADPVPVPPFGIIONNSOGYOVSTYLFRCOKY 3000  
DB 2941 VGSPPRFCCSDGTWSTGTOPSCIDPTLTTCADPVPVPPFGIIONNSOGYOVSTYLFRCOKY 3000  
QY 3001 LLOGSTTRTCLPNLWTSCTPPDCVPHRCROPETPTHANVALDLPSMGYTLITPARASP 3060

Db 3001 LLOGSTRICLPMNLTWSTGTPDCVPHCRQEPETHTANVGLDLPSNGYLTITPARASP 3060  
 Qy 3061 SRVAPSTAPARMAGQASRSAMRSGVCPSTLPGSHRSKPK 3104  
 Db 3061 SRVAPSTAPARMAGQASRSAMRSGVCPSTLPGSHRSKPK 3104  
 RESULT 5  
 ADH71136  
 ID ADH71136 standard; protein, 3546 AA.  
 AC ADH71136;  
 XX  
 DT 25-MAR-2004 (first entry)  
 XX  
 DE Human protein of the invention NOV4a SEQ ID NO:32.  
 XX  
 KM human; cytosstatic; immunomodulator; neuroprotective; nootropic;  
 KM anorectic; antidiabetic; antimicrobial; antilipidemic; gene therapy;  
 KM vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;  
 KM obesity; diabetes; infectious disease; metabolic syndrome X;  
 KM dyslipidaemia.  
 XX  
 OS Homo sapiens.  
 OS  
 PN MO200102155-A2.  
 XX  
 PD 11-DEC-2003.  
 XX  
 PF 03-JUN-2003; 2003WO-US017430.  
 XX  
 PR 03-JUN-2002; 2002US-0385120P.  
 PR 04-JUN-2002; 2002US-0385784P.  
 PR 05-JUN-2002; 2002US-0386041P.  
 PR 06-JUN-2002; 2002US-0386047P.  
 PR 06-JUN-2002; 2002US-0386376P.  
 PR 06-JUN-2002; 2002US-0386453P.  
 PR 06-JUN-2002; 2002US-0386864P.  
 PR 06-JUN-2002; 2002US-0387016P.  
 PR 07-JUN-2002; 2002US-0386796P.  
 PR 07-JUN-2002; 2002US-0386816P.  
 PR 07-JUN-2002; 2002US-0386931P.  
 PR 07-JUN-2002; 2002US-0386942P.  
 PR 07-JUN-2002; 2002US-0386971P.  
 PR 07-JUN-2002; 2002US-0387262P.  
 PR 08-JUN-2002; 2002US-0296960P.  
 PR 10-JUN-2002; 2002US-0387400P.  
 PR 10-JUN-2002; 2002US-0387535P.  
 PR 11-JUN-2002; 2002US-0387610P.  
 PR 11-JUN-2002; 2002US-0387625P.  
 PR 11-JUN-2002; 2002US-0387634P.  
 PR 11-JUN-2002; 2002US-0387668P.  
 PR 11-JUN-2002; 2002US-0387686P.  
 PR 11-JUN-2002; 2002US-0387702P.  
 PR 11-JUN-2002; 2002US-0387836P.  
 PR 11-JUN-2002; 2002US-0387859P.  
 PR 12-JUN-2002; 2002US-0387933P.  
 PR 12-JUN-2002; 2002US-0387934P.  
 PR 12-JUN-2002; 2002US-0387960P.  
 PR 12-JUN-2002; 2002US-0388022P.  
 PR 12-JUN-2002; 2002US-0388096P.  
 PR 13-JUN-2002; 2002US-0389123P.  
 PR 14-JUN-2002; 2002US-0389118P.  
 PR 14-JUN-2002; 2002US-0389120P.  
 PR 14-JUN-2002; 2002US-0389144P.  
 PR 14-JUN-2002; 2002US-0389146P.  
 PR 14-JUN-2002; 2002US-0389729P.  
 PR 17-JUN-2002; 2002US-0389742P.  
 PR 18-JUN-2002; 2002US-0389884P.  
 PR 19-JUN-2002; 2002US-0390006P.  
 PR 19-JUN-2002; 2002US-0390209P.  
 PR 21-JUN-2002; 2002US-0390763P.

PR 17-JUN-2002; 2002US-0396706P.  
 PR 06-AUG-2002; 2002US-0401628P.  
 PR 09-AUG-2002; 2002US-0402156P.  
 PR 09-AUG-2002; 2002US-0402256P.  
 PR 09-AUG-2002; 2002US-0402389P.  
 PR 12-AUG-2002; 2002US-0402786P.  
 PR 12-AUG-2002; 2002US-0402816P.  
 PR 12-AUG-2002; 2002US-0402821P.  
 PR 12-AUG-2002; 2002US-0402832P.  
 PR 13-AUG-2002; 2002US-0403448P.  
 PR 13-AUG-2002; 2002US-0403459P.  
 PR 13-AUG-2002; 2002US-0403531P.  
 PR 13-AUG-2002; 2002US-0403532P.  
 PR 13-AUG-2002; 2002US-0403563P.  
 PR 13-AUG-2002; 2002US-0406317P.  
 PR 15-AUG-2002; 2002US-0403617P.  
 PR 26-AUG-2002; 2002US-0406182P.  
 PR 26-AUG-2002; 2002US-0406355P.  
 PR 27-AUG-2002; 2002US-0406240P.  
 PR 12-SEP-2002; 2002US-0410084P.  
 PR 20-SEP-2002; 2002US-0412528P.  
 PR 23-SEP-2002; 2002US-0412731P.  
 PR 30-SEP-2002; 2002US-0414801P.  
 PR 30-SEP-2002; 2002US-0414839P.  
 PR 30-SEP-2002; 2002US-0414840P.  
 PR 30-SEP-2002; 2002US-0414954P.  
 PR 09-OCT-2002; 2002US-0417186P.  
 PR 09-OCT-2002; 2002US-0417406P.  
 PR 23-OCT-2002; 2002US-0420639P.  
 PR 28-OCT-2002; 2002US-0421156P.  
 PR 31-OCT-2002; 2002US-0422690P.  
 PR 01-NOV-2002; 2002US-0423130P.  
 PR 05-NOV-2002; 2002US-00423798.  
 PR 05-NOV-2002; 2002US-0423798P.  
 PR 12-NOV-2002; 2002US-0425453P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Alsobrook JP, Alvarez E, Anderson DM, Boldog FL, Caeman SJ,  
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K,  
 PI Ettenberg S, Gangoli EA, Gerlach VU, Gorman L, Gunther E, Guo X,  
 PI Gusev VV, Hermann JU, Ji W, Kekuda R, Li L, Liu X, MacDougall JR,  
 PI MacLachlan T, Malayanar UM, Mezick AJ, Millet I, Mehra VS,  
 PI Padigam M, Patturajan M, Pena CB, Peyman JA, Raha D, Raetelli L,  
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shimkets RA,  
 PI Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M,  
 PI Zhong H;  
 XX  
 DR WPI: 2004-081935/08.  
 DR N-PSDB; ADH71135.  
 XX  
 PT New NOVX polypeptides and nucleic acid molecules useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or  
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.  
 XX  
 PS Example 4; SEQ ID NO 32; 1880bp; English.  
 XX  
 The invention relates to a novel isolated polypeptide (NOVX). A  
 CC polypeptide of the invention has cytosstatic, immunomodulator,  
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and  
 CC antilipidemic activity, and may have a use in gene therapy, and as a  
 CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising  
 CC any of the 303 fully defined nucleotide sequences given in the  
 CC specification. The polypeptide is useful in the manufacture of a  
 CC medicament for treating a syndrome associated with a human disease. The  
 CC polypeptide, polynucleotide and antibody are useful in diagnosing.  
 CC treating or preventing NOVX-associated disorders, e.g. cancer, cachexia,  
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious  
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are  
 CC further used as hybridisation probes, in chromosome mapping, tissue  
 CC typing, preventive medicine, and pharmacogenomics. The present sequence  
 CC represents a NOVX polypeptide of the invention.

|    |   |      |
|----|---|------|
| 50 | Sequence 3546 AA:   |      |
|    | Query Match 94.5%; Score 16057; DB 8; Length 3546;                |      |
|    | Best Local Similarity 96.7%; Pred. No. 0;                         |      |
|    | Matches 2956; Conservative 14; Mismatches 49; Indels 38; Gaps 10; |      |
| QY | 2 AGAPPAL-----LPECSLISDCCASNORSHVGVPSSELVKKQIEFLKRGVLMPSKNSQ      | 57   |
| DB | 249 ASLAPAPVSSKMWLIHFTSD---GNHRQR-GFSAQIQVKKQIEFLKRGVLMPSKNSQ     | 304  |
| QY | 58 KTSVLTVGVSGQHNMCDDPGI PERGKRLGSDFLGSSVQFTCEGYDLOGSKRITCMY      | 117  |
| DB | 305 KTSVTVGVSGQHNMCDDPGI PERGKRLGSDFLGSSVQFTCEGYDLOGSKRITCMY      | 354  |
| QY | 118 SDMAASDHRPVCRAHMCDAHLRPSGIITSPNFIQYDNNAHCWIIITALLPSKIKYL      | 177  |
| DB | 365 SDMAASDHRPVCRAHMCDAHLRPSGIITSPNFIQYDNNAHCWIIITALLPSKIKYL      | 424  |
| QY | 178 AFEEFDLERGYDITLVGDGQDQDKTVLYMSQNAQSDSPHTPGSRIPESSMGDIWRQW     | 237  |
| DB | 425 AFEEFDLERGYDITLVGDGQDQDKTVLYM-----LIGTSVPLIVSTN-HQW           | 473  |
| QY | 238 TVLEICDIDISSDARSQSVKSPKTSNAVELVAPGTEIEQSGCGDPGIPAYGRREGSRH    | 297  |
| DB | 474 LLEFG-----TDG-SGS-----SLGFKASYEIEIEQSGCGDPGIPAYGRREGSRH       | 516  |
| QY | 298 HGDTLKEFCOPAFELVQKAIITCOXNNQMSAKKRGCVFSCFENFTSPSGVLSPNYEDY    | 357  |
| DB | 517 HGDTLKEFCOPAFELVQKAIITCOXNNQMSAKKRGCVFSCFENFTSPSGVLSPNYEDY    | 576  |
| QY | 358 GNHLHCWILLARPESHIHLAFNDIDIVEPODFVLIKDGTABAPVLGTFSGNQLPSIT     | 417  |
| DB | 577 GNHLHCWILLARPESHIHLAFNDIDIVEPODFVLIKDGTABAPVLGTFSGNQLPSIT     | 636  |
| QY | 418 SSGVVARLEIOTHSKRGPNITFTFRHNECPDGPVNGKRGDSILOGSIIFLGD          | 477  |
| DB | 637 SSGVVARLEIOTHSKRGPNITFTFRHNECPDGPVNGKRGDSILOGSIIFLGD          | 696  |
| QY | 478 EGFLTGSEBITTVLKEGSVVWNSAVLRCEAPCGGHLTSPSGTILSPGMFGFYKDALSC    | 537  |
| DB | 697 EGFLTGSEBITTVLKEGSVVWNSAVLRCEAPCGGHLTSPSGTILSPGMFGFYKDALSC    | 756  |
| QY | 538 AMWIEAOPGYPIKITDRFKTEVNYDTLEVNRGRITVSAPIIGVHGTQVPOFLISTNYL    | 597  |
| DB | 757 AMWIEAOPGYPIKITDRFKTEVNYDTLEVNRGRITVSAPIIGVHGTQVPOFLISTNYL    | 816  |
| QY | 598 YLLFTSDKSHSDIGFOLRYETITLOSCHLDPGI PVNGQRHGDVFVGLJNTESCDSGT    | 657  |
| DB | 817 YLLFTSDKSHSDIGFOLRYETITLOSCHLDPGI PVNGQRHGDVFVGLJNTESCDSGT    | 876  |
| QY | 658 LSGDEPLECEPNFQMSRALPSCCALCGFIOGSSGTILSPGFPDFYNNINACTWIIETSH   | 717  |
| DB | 877 LSGDEPLECEPNFQMSRALPSCCALCGFIOGSSGTILSPGFPDFYNNINACTWIIETSH   | 936  |
| QY | 718 GKGVFTPTFHEESGHVILITENGSTFQPLRLTGLSLPAPISAGLYGNFTAYORFIS      | 777  |
| DB | 937 GKGVFTPTFHEESGHVILITENGSTFQPLRLTGLSLPAPISAGLYGNFTAYORFIS      | 996  |
| QY | 778 DFSMSYEGFNITFSEYDLEPCEPEVPAYSIRKGLQFGVGDITFTSCPGFYLEGTAART    | 837  |
| DB | 997 DFSMSYEGFNITFSEYDLEPCEPEVPAYSIRKGLQFGVGDITFTSCPGFYLEGTAART    | 1056 |
| QY | 838 CUGRRRLMSSPLRVCVAECGNSVTGTGTLSPNFVNNNNHNCIYSIQTPQKGIQL        | 897  |
| DB | 1057 CUGRRRLMSSPLRVCVAECGNSVTGTGTLSPNFVNNNNHNCIYSIQTPQKGIQL       | 1116 |
| QY | 898 KARAFLESGDYLVKAYDGNNSARLIGVSHSEMMGVTLNNGSSLMDFITDANTSKG       | 957  |
| DB | 1117 KARAFLESGDYLVKAYDGNNSARLIGVSHSEMMGVTLNNGSSLMDFITDANTSKG      | 1176 |
| QY | 958 FELHFSFELLKCEDPGTKFGYKVDGSHFAGSSVSFSCDPGYSLRGSBELLCISGERR     | 1017 |
| DB | 1177 FELHFSFELLKCEDPGTKFGYKVDGSHFAGSSVSFSCDPGYSLRGSBELLCISGERR    | 1236 |

|    |   |      |
|----|---|------|
| QY | 1018 TWDRPLPTCVABCGGTIVGEVSGOVLSPGYPAPAYEHNLINCIWTIEBAECTIGLHFLVD | 1077 |
| DB | 1237 TWDRPLPTCVABCGGTIVGEVSGOVLSPGYPAPAYEHNLINCIWTIEBAECTIGLHFLVD | 1296 |
| QY | 1078 TEHVHVLIRINQPVESGVLLKELSGPALPKDLHSTFNSVLQFSTDEFTSKQFALQFS    | 1137 |
| DB | 1297 TEHVHVLIRINQPVESGVLLKELSGPALPKDLHSTFNSVLQFSTDEFTSKQFALQFS    | 1356 |
| QY | 1138 VSTATSCNDPGIPONGSREGSDMEAGDSVFQCDGVALQGAETISCVKIEHFRFMOPBP   | 1197 |
| DB | 1357 GSTATSCNDPGIPONGSREGSDMEAGDSVFQCDGVALQGAETISCVKIEHFRFMOPBP   | 1416 |
| QY | 1198 PTCIAPCGGDLTGSPGVILSPNYPBPYPGKECDMKWTVPDYLALV-FNIFNLEPGYD    | 1256 |
| DB | 1417 PTCIAPCGGDLTGSPGVILSPNYPBPYPGKECDMKWTVPDYLALV-FNIFNLEPGYD    | 1476 |
| QY | 1257 FLHITDGRDLSPLIGSFYGSQLPGRITSSSNSLFLARSDASVSNAGFVIDFENPREB    | 1316 |
| DB | 1477 FLHITDGRDLSPLIGSFYGSQLPGRITSSSNSLFLARSDASVSNAGFVIDFENPREB    | 1536 |
| QY | 1317 CFDPGSIKNGTRVSGDLKLGSSVTYYCHGAYEVEGTSTLSCILGPDGKPVNNNPVCTA   | 1376 |
| DB | 1537 CFDPGSIKNGTRVSGDLKLGSSVTYYCHGAYEVEGTSTLSCILGPDGKPVNNNPVCTA   | 1596 |
| QY | 1377 PCGGQYVSDGVYLSPNYPONTSGQICLYFVTVPKDY-VVFGQPAFHTALNDVEVHD     | 1435 |
| DB | 1597 PCGGQYVSDGVYLSPNYPONTSGQICLYFVTVPKDYVFGQPAFHTALNDVEVHD       | 1656 |
| QY | 1436 GHISCHRLSSLSGSHTESLPLATSNQVLKFSKGLAPARGHFPYQVAPRTSATQCS      | 1495 |
| DB | 1657 GHISCHRLSSLSGSHTESLPLATSNQVLKFSKGLAPARGHFPYQVAPRTSATQCS      | 1716 |
| QY | 1496 SVPEPRYKRLGSDPSVGAIVRFECSNGYALQGSPEIECLFVPALAQMNVSAPTCVPC    | 1555 |
| DB | 1717 SVPEPRYKRLGSDPSVGAIVRFECSNGYALQGSPEIECLFVPALAQMNVSAPTCVPC    | 1776 |
| QY | 1556 GGNLTERRGTLISGFPFPYNSLNCWKIVPFGAGIQIOVVSFVTEQWMDSEVDDGA      | 1615 |
| DB | 1777 GGNLTERRGTLISGFPFPYNSLNCWKIVPFGAGIQIOVVSFVTEQWMDSEVDDGA      | 1836 |
| QY | 1616 DNTVYMLGSFSGITVPALLNSTSNQLYLHFSYDSVSAAGFHLKXKTVGSSCEPEAVPS   | 1675 |
| DB | 1837 DNTVYMLGSFSGITVPALLNSTSNQLYLHFSYDSVSAAGFHLKXKTVGSSCEPEAVPS   | 1896 |
| QY | 1676 NGVKTGERYLVNDVVSQCEBEGVALQGHAIISCMPTGVRRANYPPLICIAQCGGTEBME  | 1735 |
| DB | 1897 NGVKTGERYLVNDVVSQCEBEGVALQGHAIISCMPTGVRRANYPPLICIAQCGGTEBME  | 1956 |
| QY | 1736 GVILSPGPPGNYPSNMCMCKIALPVGFGAHIOFLNSTEPMNDYIEIRNGPYETSRMG    | 1795 |
| DB | 1957 GVILSPGPPGNYPSNMCMCKIALPVGFGAHIOFLNSTEPMNDYIEIRNGPYETSRMG    | 2016 |
| QY | 1796 RFSGSELPSLSLSTSHETTYFPHSDHQNRPGRFLXQVAVLEQCDPBEFANGIVAGAG    | 1855 |
| DB | 2017 RFSGSELPSLSLSTSHETTYFPHSDHQNRPGRFLXQVAVLEQCDPBEFANGIVAGAG    | 2076 |
| QY | 1856 YNNGOSVTEFCLEPQYOLTGHPVLTQCHGTRNMDHPLPKCEVRCGNTTSSNGTVSPGF   | 1915 |
| DB | 2077 YNNGOSVTEFCLEPQYOLTGHPVLTQCHGTRNMDHPLPKCEVRCGNTTSSNGTVSPGF   | 2136 |
| QY | 1916 PSFYSSQDCVWLITVYIGHGVRNLNLSLQTEBESGDIITIMDPOOQAPAPLGVFTSMK   | 1975 |
| DB | 2137 PSFYSSQDCVWLITVYIGHGVRNLNLSLQTEBESGDIITIMDPOOQAPAPLGVFTSMK   | 2196 |
| QY | 1976 KTYOSSNOVLKFRHDAATGGAIFALAFSAYPLTKCPPTLIPNAEVTENEBFNIGDIY    | 2035 |
| DB | 2197 KTYOSSNOVLKFRHDAATGGAIFALAFSAYPLTKCPPTLIPNAEVTENEBFNIGDIY    | 2256 |
| QY | 2036 RYRCLRGFTLVGNELLTKLGTVLOEGRPPICEVHCPTNELLTDSGTVILSGSYGSGYP   | 2095 |
| DB | 2257 RYRCLRGFTLVGNELLTKLGTVLOEGRPPICEVHCPTNELLTDSGTVILSGSYGSGYP   | 2316 |



PR 26-AUG-2002; 2002US-0406182P.  
 PR 26-AUG-2002; 2002US-0406355P.  
 PR 27-AUG-2002; 2002US-0406240P.  
 PR 12-SEP-2002; 2002US-0410084P.  
 PR 20-SEP-2002; 2002US-0412528P.  
 PR 23-SEP-2002; 2002US-0412731P.  
 PR 30-SEP-2002; 2002US-0414801P.  
 PR 30-SEP-2002; 2002US-0414839P.  
 PR 30-SEP-2002; 2002US-0414840P.  
 PR 30-SEP-2002; 2002US-0414954P.  
 PR 09-OCT-2002; 2002US-0417186P.  
 PR 23-OCT-2002; 2002US-0417406P.  
 PR 28-OCT-2002; 2002US-0420639P.  
 PR 31-OCT-2002; 2002US-0422690P.  
 PR 01-NOV-2002; 2002US-0423130P.  
 PR 05-NOV-2002; 2002US-0423798P.  
 PR 12-NOV-2002; 2002US-0425453P.  
 PR (CURA-) CURAGEN CORP.  
 XX  
 PA  
 PI Albrock JP, Alvarez E, Anderson DM, Boldog FI, Caeman SJ, Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K, Etenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X, Guev VY, Hermann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR, MacLachlan T, Malyanekar UM, Mexick AJ, Millet I, Mishra VS, Pacteraj M, Paturajan M, Pena CE, Peyman JA, Raha D, Rastelli L, Rieger DK, Rothenberg ME, Sclere P, Shenoy SG, Shinkels RA, Smithson G, Spytek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M, Zhong H;  
 PI  
 PI WPI; 2004-081935/08.  
 DR N-PSDB; ADH71143.  
 XX  
 XX New NOXV polypeptides and nucleic acid molecules useful for preventing or treating NOXV-associated disorders, e.g. cancer, diabetes, infection or obesity, and in chromosome mapping, tissue typing or pharmacogenomics.  
 PT  
 PT  
 PT  
 XX  
 PS Example 4; SEQ ID NO 40; 1880bp; English.  
 XX  
 CC The invention relates to a novel isolated polypeptide (NOXV). A  
 CC polypeptide of the invention has cytostatic, immunomodulator, and neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and antihypertensive activity, and may have a use in gene therapy, and as a vaccine. The polypeptides are encoded by NOXV polynucleotides comprising any of the 303 fully defined nucleotide sequences given in the CC specification. The polypeptide is useful in the manufacture of a CC medicament for treating a syndrome associated with a human disease. The CC polypeptide, polynucleotide and antibody are useful in diagnosing, CC treating or preventing NOXV-associated disorders, e.g. cancer, cachexia, Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are CC further used as hybridisation probes, in chromosome mapping, tissue CC typing, preventive medicine, and pharmacogenomics. The present sequence represents a NOXV polypeptide of the invention.  
 CC  
 CC  
 XX  
 SO Sequence 3483 AA;  
 Query Match 92.2%; Score 15655.5; DB 8; Length 3483;  
 Best Local Similarity 94.6%; Pred. No. 0;  
 Matches 2893; Conservative 14; Mismatches 49; Indels 101; Gaps 11;  
 QY 2 AAGPPAL-----LPPSLISDCCASQNRHSVGVSELVKQIELKSRGVKLMPKDNQ 57  
 DB 249 ASLPAAVSSIKWMLRLHFTSD--GNHROR-GFSAQYQVKQIELKSRGVKLMPKDNQ 304  
 QY 58 KTSVTLQVGVSOCHNCPDGGI PERGKRGSDPRGSSVQTFNCGYLOQSKRTTCMKV 117  
 DB 305 KTSVTLQVGVSOCHNCPDGGI PERGKRGSDPRGSSVQTFNCGYLOQSKRTTCMKV 364  
 QY 118 SMFPAAMSDHPRVRCARMCDALRGPSGIITSPPNPIQYDNNAHCWMTITLNPSTKYL 177

DB 365 SMFPAAMSDHPRVRCARMCDALRGPSGIITSPPNPIQYDNNAHCWMTITLNPSTKYL 424  
 QY 178 APEEPDLERGYDTLTVGDDGDDGDDQKTVLYMGQNCASDSEPHTPGSRIPESMGDIWROKM 237  
 DB 425 APEEPDLERGYDTLTVGDDGDDGDDQKTVLYM-----LTGTSVPLIYSTN-HQNM 473  
 QY 238 TVLEICRDISSDANSGSVRKSPKTSNAVELVAPGTELEQSGCGDPGIPAYRRREGSRPH 297  
 DB 474 LLEQ-----TDG-SGS-----SLGFASYSYEEIEQSGCGDPGIPAYRRREGSRPH 516  
 QY 298 HSDTLKFCOPAFELVYGOKAITCOKNOMSAKPCGVSCFPNFTSPGCVLSPVYREY 357  
 DB 517 HSDTLKFCOPAFELVYGOKAITCOKNOMSAKPCGVSCFPNFTSPGCVLSPVYREY 576  
 QY 358 GNLHCWMLILARPSRILHAFNDIDVBPQDFLYIKDGTAEAVLGTSPSGNQPLSSIT 417  
 DB 577 GNLHCWMLILARPSRILHAFNDIDVBPQDFLYIKDGTAEAVLGTSPSGNQPLSSIT 636  
 QY 418 SSGHVARLEFQDHSHTGKRGFNITTTFRNHCNCPGVIPVNGKRGDSLQJGSSISFLCD 477  
 DB 637 SSGHVARLEFQDHSHTGKRGFNITTTFRNHCNCPGVIPVNGKRGDSLQJGSSISFLCD 696  
 QY 478 EGFLEGTQSEETITVLAKEGSVVMNSAVLRCEAPCGHLLTSPSGTILSPGWGPFYDALSC 537  
 DB 697 EGFLEGTQSEETITVLAKEGSVVMNSAVLRCEAPCGHLLTSPSGTILSPGWGPFYDALSC 756  
 QY 538 AMVIAQGVPIKITEPRPKTEVNVDTLEVDRGRTYSAPLIGYHGTQVPOPLISTSNYL 597  
 DB 757 AMVIAQGVPIKITEPRPKTEVNVDTLEVDRGRTYSAPLIGYHGTQVPOPLISTSNYL 816  
 QY 598 YLLFSTDSSHSDIGOLAEYETITLQSDHCIDPGI PVNGQRHGNFVYCATLTFSCDSGYT 657  
 DB 817 YLLFSTDSSHSDIGOLAEYETITLQSDHCIDPGI PVNGQRHGNFVYCATLTFSCDSGYT 876  
 QY 817 YLLFSTDSSHSDIGOLAEYETITLQSDHCIDPGI PVNGQRHGNFVYCATLTFSCDSGYT 876  
 DB 877 LSDGEPLCEBNFQMSRALPSCFALCGGFIQSSGTLISPPFPFYNLNCTWIIETSH 936  
 QY 718 GKGVFFTHHTLSSGHDYLLITENGSTFQRLQTLGSLRLPAPISAGLYGNFTAQVRPIS 777  
 DB 937 GKGVFFTHHTLSSGHDYLLITENGSTFQRLQTLGSLRLPAPISAGLYGNFTAQVRPIS 996  
 QY 778 DFSMSYEGFNITFSEYDLEPCSEPEVPAVSIIRKLOQGVGDTLTFSCFPGRLEGTRIT 837  
 DB 997 DFSMSYEGFNITFSEYDLEPCSEPEVPAVSIIRKLOQGVGDTLTFSCFPGRLEGTRIT 1056  
 QY 838 CLGRRRLMSPLPRCAECGNSVTGTGTLISPNFPVNNNHCTIYSIQTOGKGIOL 897  
 DB 1057 CLGRRRLMSPLPRCAECGNSVTGTGTLISPNFPVNNNHCTIYSIQTOGKGIOL 1116  
 QY 898 KARAFELSEGVLKVYQGNNSARLIGVFSHEMKGVTLNSTSSLMLDITTAENTSKG 957  
 DB 1117 KARAFELSEGVLKVYQGNNSARLIGVFSHEMKGVTLNSTSSLMLDITTAENTSKG 1176  
 QY 958 FELHFSFELIKCEDPPTPKFGYKVHDEGHFAGSSVSFSCDPGYSLRGSBELCLSGERR 1017  
 DB 1177 FELHFSFELIKCEDPPTPKFGYKVHDEGHFAGSSVSFSCDPGYSLRGSBELCLSGERR 1236  
 QY 1018 TWDRPLPTCAECCGTVRGEVSGOVLSPGYPAEYHNLCIWTIEAAGCTTGLHFLVFD 1077  
 DB 1237 TWDRPLPTCAECCGTVRGEVSGOVLSPGYPAEYHNLCIWTIEAAGCTTGLHFLVFD 1296  
 QY 1078 TEEVHDVLRIMDPVRESGVLLKELSGPALPKDLSHFNSVVLQFSTDPFSSKGFALQPS 1137  
 DB 1297 TEEVHDVLRIMDPVRESGVLLKELSGPALPKDLSHFNSVVLQFSTDPFSSKGFALQPS 1356  
 QY 1138 VSTATSCNDPGLPONGSRSGDSWEAGDSIVFOCDPGVALQGSALISCVKIEENFFWQPS 1197  
 DB 1357 VSTATSCNDPGLPONGSRSGDSWEAGDSIVFOCDPGVALQGSALISCVKIEENFFWQPS 1416  
 QY 1198 PTCIAPCGDLTPSGVILLSPNPEYPPGKEDCMKVTPDVIALVLFPSENLPGYD 1256  
 DB 1417 PTCIAPCGDLTPSGVILLSPNPEYPPGKEDCMKVTPDVIALVLFPSENLPGYD 1476

|    |      |  |      |
|----|------|--|------|
| Qy | 1257 | FLHIDGRDLSPLIGSFYQSLPGRTESSNSLFLAFRSDASNNAGFVTDYENRES      | 1316 |
| Dd | 1477 | FLHIDGRDLSPLIGSFYQSLPGRTESSNSLFLAFRSDASNNAGFVTDYENRES      | 1536 |
| Qy | 1317 | CFDPSIKNGTRVSDKLGSVTVYCHGVEVGTSLSCILGPDGKPVNNRPVCTA        | 1376 |
| Dd | 1537 | CFDPSIKNGTRVSDKLGSVTVYCHGVEVGTSLSCILGPDGKPVNNRPVCTA        | 1596 |
| Qy | 1377 | PCGGQYVSGDGVLSPPYPNQYTSQICLIFVTVPKDY-VVFGQPAFPTALNDVEVD    | 1435 |
| Dd | 1597 | PCGGQYVSGDGVLSPPYPNQYTSQICLIFVTVPKDYVFGQPAFPTALNDVEVD      | 1656 |
| Qy | 1436 | GHSQHSRLSLSSGSHGESLPLATSNQVLIKFSAKGLAPRGHFYQVAPRTSAQCS     | 1495 |
| Dd | 1657 | GHSQHSRLSLSSGSHGESLPLATSNQVLIKFSAKGLAPRGHFYQVAPRTSAQCS     | 1716 |
| Qy | 1496 | SVEPRYXKRLGSDFSVCAIVREFCNSGYALQGSPEIECLPVGALAQMNVSAFTCVPC  | 1555 |
| Dd | 1717 | SVEPRYXKRLGSDFSVCAIVREFCNSGYALQGSPEIECLPVGALAQMNVSAFTCVPC  | 1776 |
| Qy | 1556 | GNUTERRGTLISGFPEPYLNSLNCWKIVPEAGIQIOVVSFVTEQWMSLIEVFGA     | 1615 |
| Dd | 1777 | GNUTERRGTLISGFPEPYLNSLNCWKIVPEAGIQIOVVSFVTEQWMSLIEVFGA     | 1836 |
| Qy | 1616 | DNTVTMLGSPGCTVPALLNSTSNQYLHFPYSDISVAGAFHLEKTVGLSSCEPAVPS   | 1675 |
| Dd | 1837 | DNTVTMLGSPGCTVPALLNSTSNQYLHFPYSDISVAGAFHLEKTVGLSSCEPAVPS   | 1896 |
| Qy | 1676 | NGVATGERVYLVNDVVSFCQCEGYALQGHAIHSCMPGTVRNNYPPCLIAQCGTVEME  | 1735 |
| Dd | 1897 | NGVATGERVYLVNDVVSFCQCEGYALQGHAIHSCMPGTVRNNYPPCLIAQCGTVEME  | 1956 |
| Qy | 1736 | GVLISPEPPGNYPSNMDCSWKIALPVGFGAHIQFLNFSTEBNHDYIEIRNGPYETSMG | 1795 |
| Dd | 1957 | GVLISPEPPGNYPSNMDCSWKIALPVGFGAHIQFLNFSTEBNHDYIEIRNGPYETSMG | 2016 |
| Qy | 1796 | RFSGSELPSSLLSHETTVFHSDHSONRGFLXEOAYELOCPDEBPANIVGAG        | 1855 |
| Dd | 2017 | RFSGSELPSSLLSHETTVFHSDHSONRGFLXEOAYELOCPDEBPANIVGAG        | 2076 |
| Qy | 1856 | YNNQGSVTFPECLPQYLTGHPVLTCQHTNRNMHPKCEVPCGAGITSSNGTVSPGF    | 1915 |
| Dd | 2077 | YNNQGSVTFPECLPQYLTGHPVLTCQHTNRNMHPKCEVPCGAGITSSNGTVSPGF    | 2136 |
| Qy | 1916 | PSPPYSSQDCVWLITVPDIGHGVRLNLSLQTEPSGDFITTWGQQTAPRLGVFTSMAX  | 1975 |
| Dd | 2137 | PSPPYSSQDCVWLITVPDIGHGVRLNLSLQTEPSGDFITTWGQQTAPRLGVFTSMAX  | 2196 |
| Qy | 1976 | KTVQSSNOVLLKTHRDAAATGIFAIAFSAVPLKCCPPTLLPMAEVTEENEFNIDIV   | 2035 |
| Dd | 2197 | KTVQSSNOVLLKTHRDAAATGIFAIAFSAVPLKCCPPTLLPMAEVTEENEFNIDIV   | 2256 |
| Qy | 2036 | RYRCLPFTLVGNELITCKLTQYLOFGAPPRICEVHCPTEMLLTDSTGVILSOSYPSYP | 2095 |
| Dd | 2257 | RYRCLPFTLVGNELITCKLTQYLOFGAPPRICEVHCPTEMLLTDSTGVILSOSYPSYP | 2316 |
| Qy | 2096 | QFQTCMWLVREPDYNI SLIVYFLSKQYDEFEIPDPSQGSPLKALSGNYSAPLIYT   | 2155 |
| Dd | 2317 | QFQTCMWLVREPDYNI SLIVYFLSKQYDEFEIPDPSQGSPLKALSGNYSAPLIYT   | 2376 |
| Qy | 2156 | SSNSVYLRMSDDHANNRKGFKIRYSAPYCSLPPAALHGLTQSTQPSGSHFGCMAG    | 2215 |
| Dd | 2377 | SSNSVYLRMSDDHANNRKGFKIRYSAPYCSLPPAALHGLTQSTQPSGSHFGCMAG    | 2402 |
| Qy | 2216 | YRLVGHSMALCTRRHPOGHLWSEAIPLCOALSGLPEAPKNGVFGKEYTVGTAAVYSCS | 2275 |
| Dd | 2403 | YRLVGHSMALCTRRHPOGHLWSEAIPLCOALSGLPEAPKNGVFGKEYTVGTAAVYSCS | 2443 |
| Qy | 2276 | EGYHLOAGAEATAECLDTGLMSNNRNPQCVPTCPDVSSISVEHGMRLIFETOYQFOA  | 2335 |
| Dd | 2434 | EGYHLOAGAEATAECLDTGLMSNNRNPQCVPTCPDVSSISVEHGMRLIFETOYQFOA  | 2493 |

  

|    |      |   |      |
|----|------|---|------|
| Qy | 2336 | QLMILCDPGYYTGORVIRCOANGKMSLGDSTPTCRILISGCELPPIPNHGRICTLASYGA  | 2395 |
| Dd | 2494 | QLMILCDPGYYTGORVIRCOANGKMSLGDSTPTCRILISGCELPPIPNHGRICTLASYGA  | 2553 |
| Qy | 2396 | TAIFSCNSGYTLVGSRVRECMANGWLSGSEVRCLAGHCGTPEPIVNGHNGENYSYRGV    | 2455 |
| Dd | 2554 | TAIFSCNSGYTLVGSRVRECMANGWLSGSEVRCLAGHCGTPEPIVNGHNGENYSYRGV    | 2613 |
| Qy | 2456 | VYQCNAGFLIGMSVRI COODHMSGKTPFCVITITGHPENPNVGLTQGNQFLNDVYKF    | 2515 |
| Dd | 2614 | VYQCNAGFLIGMSVRI COODHMSGKTPFCVITITGHPENPNVGLTQGNQFLNDVYKF    | 2673 |
| Qy | 2516 | VCPNPGVABEAGARSOCIASGQMSDMLPTCPI INCTPBGHQENSVRQVHAGSPRFSFGTT | 2575 |
| Dd | 2674 | VCPNPGVABEAGARSOCIASGQMSDMLPTCPI INCTPBGHQENSVRQVHAGSPRFSFGTT | 2733 |
| Qy | 2576 | VSYRCHNGFYLLGTPLYLSCQGDGTWDRPREOCLVSCGHPSPHSQMSGDSYTVGAVVR    | 2635 |
| Dd | 2734 | VSYRCHNGFYLLGTPLYLSCQGDGTWDRPREOCLVSCGHPSPHSQMSGDSYTVGAVVR    | 2793 |
| Qy | 2636 | YSCIGKRTLVGNSTRMCGLDGHWTSGLPHSCSGTSVYCGDPGIPAHGIRLGDSPDPTVM   | 2695 |
| Dd | 2794 | YSCIGKRTLVGNSTRMCGLDGHWTSGLPHSCSGTSVYCGDPGIPAHGIRLGDSPDPTVM   | 2853 |
| Qy | 2696 | RFSCEAGHYLRGSSERTCOANGSMGSOPECGVYISGPNPTPSNARVVPFDGLVFSSSIY   | 2755 |
| Dd | 2854 | RFSCEAGHYLRGSSERTCOANGSMGSOPECGVYISGPNPTPSNARVVPFDGLVFSSSIY   | 2913 |
| Qy | 2756 | YECREGYATGLSRHCSVNGVTGSDPECLVINCQDPGIPANGRLRGDNFRYNTVYQ       | 2815 |
| Dd | 2914 | YECREGYATGLSRHCSVNGVTGSDPECLVINCQDPGIPANGRLRGDNFRYNTVYQ       | 2973 |
| Qy | 2816 | CVPGYMESHRSVYLSCTQRTNNGTKPVCKALMKCPPLIPNGKYVGSPPMGSSVTVYA     | 2875 |
| Dd | 2974 | CVPGYMESHRSVYLSCTQRTNNGTKPVCKALMKCPPLIPNGKYVGSPPMGSSVTVYA     | 3033 |
| Qy | 2876 | CLBGYOLSLPAVFTCEGNSWTELPQCFVFCGDPGVPGRGRREDGFSYRSVSFSCH       | 2935 |
| Dd | 3034 | CLBGYOLSLPAVFTCEGNSWTELPQCFVFCGDPGVPGRGRREDGFSYRSVSFSCH       | 3093 |
| Qy | 2936 | PPLVLVGSPPRRFCOSDGTWSTGTPSCIDPILTTCADPGVBOFGIIONNSQGYOVSIVLFR | 2995 |
| Dd | 3094 | PPLVLVGSPPRRFCOSDGTWSTGTPSCIDPILTTCADPGVBOFGIIONNSQGYOVSIVLFR | 3153 |
| Qy | 2996 | COKGYLLQGSTTTCPLPNTLWSTGTPDPCVPHHROBETPTHANVAGALDPSMGYTLI     | 3052 |
| Dd | 3154 | COKGYLLQGSTTTCPLPNTLWSTGTPDPCVPHHROBETPTHANVAGALDPSMGYTLI     | 3210 |

  

|          |  |  |  |
|----------|--|--|--|
| RESULT 7 |  |  |  |
| ADH71146 | ID ADH71146 standard; protein; 3130 AA.                              |  |  |
| XX       | XX   |  |  |
| AC       | ADH71146;  |  |  |
| XX       | XX   |  |  |
| DT       | 25-MAR-2004 (first entry)  |  |  |
| XX       | XX   |  |  |
| DE       | Human protein of the invention NOVAE SEQ ID NO:42.                   |  |  |
| XX       | XX   |  |  |
| KW       | human; cytosolic; immunomodulator; neuroprotective; nootropic;       |  |  |
| KW       | anorectic; antidiabetic; antimicrobial; antipneumonia; gene therapy; |  |  |
| KW       | vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease; |  |  |
| KW       | obesity; diabetes; infectious disease; metabolic syndrome X;         |  |  |
| KW       | dyslipidemia.  |  |  |
| XX       | XX   |  |  |
| OS       | Homo sapiens.  |  |  |
| XX       | XX   |  |  |
| PN       | NO2003102155-A2.   |  |  |
| XX       | XX   |  |  |
| PD       | 11-DEC-2003.   |  |  |
| XX       | XX   |  |  |
| PF       | 03-JUN-2003; 2003WO-US017430.  |  |  |
| XX       | XX   |  |  |

|    |   |   |   |
|----|---|---|---|
|    | PA  | (CURA-) CURAGEN CORP.   |   |
|    | XX  |   |   |
| Pt | PI  | Aleobrook JP, Alvarez E, Anderson DW, Boldog FI, Casman SJ,               |   |
| DR | PI  | Catterton E, Chapoval A, Crabtree-Boker JR, Edinger SR, Ellerman K,       |   |
| XX | PI  | Ettenton S, Gangoli EA, Getlach VL, Gotman L, Gunther E, Guo X,           |   |
| XX | PI  | Gusev VI, Herrmann JL, Ji W, Kakuda R, Li L, Liu X, MacDougall JR,        |   |
| PI | PI  | Machlan T, Malyanar UM, Mezik AJ, Millet I, Mishra VS,                    |   |
| PI | PI  | Padiagaran M, Paturajan M, Pena CE, Peyman JA, Raha D, Raetelli L,        |   |
| PI | PI  | Rieger DK, Rothenberg ME, Score P, Shenoy SG, Shinkels RA,                |   |
| PI | PI  | Smlthson G, Speytek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M,           |   |
| XX | PI  | Zhong H;  |   |
| XX |   |   |   |
| DR | N-PSDB; ADH71145.   | WPI; 2004-081935/08.  |   |
| PT |   |   |   |
| PT |   | New NOXV polypeptides and nucleic acid molecules useful for preventing or |   |
| PT |   | treating NOXV-associated disorders, e.g. cancer, diabetes, infection or   |   |
| PT |   | obesity, and in chromosome mapping, tissue typing or pharmacogenomics.    |   |
| XX |   |   |   |
| XX |   | Example 4; SEQ ID NO 42; 1880bp; English.                                 |   |
| CC |   | The invention relates to a novel isolated polypeptide (NOXV). A           |   |
| CC |   | polypeptide of the invention has cytostatic, immunomodulator,             |   |
| CC |   | neuroprotective, nocotropic, anorectic, antidiabetic, antimicrobial, and  |   |
| CC |   | antihypaemic activity, and may have a use in gene therapy, and as a       |   |
| CC |   | vaccine. The polypeptides are encoded by NOXV polynucleotides comprising  |   |
| CC |   | any of the 303 fully defined nucleotide sequences given in the            |   |
| CC |   | specification. The polypeptide is useful in the manufacture of a          |   |
| CC |   | medicament for treating a syndrome associated with a human disease. The   |   |
| CC |   | polypeptide, polynucleotide and antibody are useful in diagnosing,        |   |
| CC |   | treating or preventing NOXV-associated disorders, e.g. cancer, cachexia,  |   |
| CC |   | Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious   |   |
| CC |   | diseases, metabolic syndrome X or dyslipidemias. The nucleic acids are    |   |
| CC |   | further used as hybridisation probes, in chromosome mapping, tissue       |   |
| CC |   | typing, preventive medicine, and pharmacogenomics. The present sequence   |   |
| CC |   | represents a NOXV polypeptide of the invention.                           |   |
| XX |   |   |   |
| SO |   | Sequence 3130 AA;   |   |
|    |   |   |   |
|    | Query Match   | 89.6%; Score 15219; DB 8; Length 3130;                                    |   |
|    | Best Local Similarity   | 93.5%; Pred. No. 0;   |   |
|    | Matches 2812; Conservative  | 13; Mismatches 29; Indels 152; Gaps                                       | 9 |
| QY | 50 MSKNSQSTSVLTVOYGSGCHNCCDPPIGPERKGRLGSFRLGSSVOFTCNEGYLGS        | 109   |   |
| Db | 1 MSKNSQSTSVTVOYGSGCHNCCDPPIGPERKRGISFRLGSSVOFTCNEGYDLS           | 60  |   |
| QY | 110 KRITCMKSDMFAAMSDIRPCARMCDALRGSGLIISPSPFIQDNNAHCVTLTAL         | 169   |   |
| Db | 61 KRITCMKSDMFAAMSDIRPCARMCDALRGSGLIISPSPFIQDNNAHCVTLTAL          | 120   |   |
| QY | 170 NBSKVILAFEEFDLERGYDTLTVGDGGODQDKTYLYMSONACSDSPHTPGSRIPESMS    | 229   |   |
| Db | 121 NBSKVILAFEEFDLERGYDTLTVGDGGODQDKTYLYM-----LTGVSVDLLIV         | 170   |   |
| QY | 220 GDIMOKVTLEICRDJSSDGABSGSVKPSKTSNAVELVAPEETIEOGSCDDPGIPAYG     | 289   |   |
| Db | 171 STM-HQMMLLPQ-----TDG-SGS-----SLGFRAVEEIEOGSCDDPGIPAYG         | 212   |   |
| QY | 290 RREGSRFHAGDTLTFECOPAFELVGQAITYCQKNOMSAKKGCQVSPCFENFTSPSGVL    | 349   |   |
| Db | 213 RREGSRFHAGDTLTFECOPAFELVGQAITYCQKNOMSAKKGCQVSPCFENFTSPSGVL    | 272   |   |
| QY | 350 SPNPEDYGNHLHCVMILLARPSRIHLAENDIDIVEPODFLVIXDGAATAEARPVLGTFSG  | 409   |   |
| Db | 273 SPNPVEDYGNHLHCVMILLARPSRIHLAENDIDIVEPODFLVIXDGAATAEARPVLGTFSG | 332   |   |
| QY | 410 NOLBSTSSGHVARLEOFNDHGTGRGRNTPTTTFRNHAGCDPGVPVNGKRFQSLQIG      | 469   |   |

|    |      |   |      |
|----|------|---|------|
| Db | 333  | NOLPSSITSSGHWARLEFOTDHTSGKRGFNITTSSESSNECPDGPVFNKRGFGDSLQ     | 392  |
| Qy | 470  | SSISFLDDEGLTQGSSTITTCVUKESGVNNSAVLKEAPCGHLLSPSSTILSPGMP       | 529  |
| Db | 393  | SSISFLDDEGLTQGSSTITTCVUKESGVNNSAVLKEAPCGHLLSPSSTILSPGMP       | 452  |
| Qy | 530  | PKYALSCAMVIEAOPGPIKITFDRPKTEVNYDTLEVRDGRYSAPLIGYHGTQVPOF      | 589  |
| Db | 453  | PKYALSCAMVIEAOPGPIKITFDRPKTEVNYDTLEVRDGRYSAPLIGYHGTQVPOF      | 512  |
| Qy | 590  | LISTSNTLYLLFSTDKSHDIGFOLRYETTLQSDHCLDPGI PVNGQRHNDFFVGLAVT    | 649  |
| Db | 513  | LISTSNTLYLLFSTDKSHDIGFOLRYEATLQSDHCLDPGI PVNGQRHNDFFVGLAVT    | 572  |
| Qy | 650  | FSCSGTTLSDGELECEPNFQMSRALPSCCALCGGFI QGSSGTTILSPGPFDPYNNLNC   | 709  |
| Db | 573  | FSCSGTTLSDGELECEPNFQMSRALPSCCALCGGFI QGSSGTTILSPGPFDPYNNLNC   | 632  |
| Qy | 710  | TWIIETSHGKGVFTFHTFHLESCHDYLLITENGSPFOPLRQLTGSRLPAPISAGLYNF    | 769  |
| Db | 633  | TWIIETSHGKGVFTFHTFHLESCHDYLLITENGSPFOPLRQLTGSRLPAPISAGLYNF    | 692  |
| Qy | 770  | TAQVRPISDFMSYEGFNITFSEYDLBECPEVPAYSIRKGLQFGVGDTLTFSGCPGYR     | 829  |
| Db | 693  | TAQVRPISDFMSYEGFNITFSEYDLBECPEVPAYSIRKGLQFGVGDTLTFSGCPGYR     | 752  |
| Qy | 830  | LEGARATTCGGRRLWSSPLPRCAECGNSVTGTQGLTSPN PVNNTNNHECTIYSIQ      | 889  |
| Db | 753  | LEGARATTCGGRRLWSSPLPRCAECGNSVTGTQGLTSPN PVNNTNNHECTIYSIQ      | 812  |
| Qy | 890  | QPGKGIOLKARAFELSEGDVLYKVDGNNNSARLLGVF SHSEMVMVTLNSTSSILMLFIT  | 949  |
| Db | 813  | QPGKGIOLKARAFELSEGDVLYKVDGNNNSARLLGVF SHSEMVMVTLNSTSSILMLFIT  | 872  |
| Qy | 950  | DAENTSKGFBELHFSFBLIKCEDPGTPKFGYKVDHGHFAGSSVFSFGCDPGYSLRGSEL   | 1009 |
| Db | 873  | DAENTSKGFBELHFSFBLIKCEDPGTPKFGYKVDHGHFAGSSVFSFGCDPGYSLRGSEL   | 932  |
| Qy | 1010 | LCLSGERRTMDRPLPTCYAECGTVRGEVSGQVLSFGYPRAPYENHNLCTITTEAGCTI    | 1069 |
| Db | 933  | LCLSGERRTMDRPLPTCYAECGTVRGEVSGQVLSFGYPRAPYENHNLCTITTEAGCTI    | 992  |
| Qy | 1070 | GLHFLVDTBEVHVLRIWDPVPSGVLLKELSGPALPKDLHSTFNSVVLQFSTDPFTSK     | 1129 |
| Db | 993  | GLHFLVDTBEVHVLRIWDPVPSGVLLKELSGPALPKDLHSTFNSVVLQFSTDPFTSK     | 1062 |
| Qy | 1130 | QGFALIOFSVSTATSCNDPGIPONGSRSGDSWEAGDSTVFQCDPGYALOGSAEISCXYEN  | 1189 |
| Db | 1053 | QGFALIOFSVSTATSCNDPGIPONGSRSGDSWEAGDSTVFQCDPGYALOGSAEISCXYEN  | 1112 |
| Qy | 1190 | RFFMQPSBPPTCIAPCGGDLTGPSSVILSPNYPEPYPEGKECDMKVTVSPDYVALV-FNI  | 1248 |
| Db | 1113 | RFFMQPSBPPTCIAPCGGDLTGPSSVILSPNYPEPYPEGKECDMKVTVSPDYVALV-FNI  | 1172 |
| Qy | 1249 | FNLBPYDHLIYVGRDLSPLISFYGSOLRGRIEISSNSLFLAFPSDASVSNAAGVTD      | 1308 |
| Db | 1173 | FNLBPYDHLIYVGRDLSPLISFYGSOLPGRIEISSNSLFLAFPSDASVSNAAGVTD      | 1232 |
| Qy | 1309 | YTERPRESCEPDGSIKNGTRVSGDLKLGSSVTVYCHGGEVEGSTRSCLIGPDGKPVWN    | 1368 |
| Db | 1233 | YTERPRESCEPDGSIKNGTRVSGDLKLGSSVTVYCHGGEVEGSTRSCLIGPDGKPVWN    | 1292 |
| Qy | 1369 | NPRPVCTAPCGGQYVSGDVLSPNYPONTYSSGOI CLYFTVTPYQY-VVFGQPAFFHTAL  | 1427 |
| Db | 1293 | NPRPVCTAPCGGQYVSGDVLSPNYPONTYSSGOI CLYFTVTPYQY-VVFGQPAFFHTAL  | 1352 |
| Qy | 1428 | NDVVEVMDHSHOHRLLSSLSGSHT-GBSLPLATSNQVLIKFSKAGLAPARGPFPVQAV    | 1486 |
| Db | 1353 | NDVVEVMDHSHOHRLLSSLSGSHTGBSLPLATSNQVLIKFSKAGLAPARGPFPVQAV     | 1412 |
| Qy | 1487 | PRTSATOCSSVPEBRYGKRLGSDPSVGAI VRFECNSGVALOGSPEIECLPVGALAOMNV  | 1546 |
| Db | 1413 | PRTSATOCSSVPEBRYGKRLGSDPSVGAI VRFECNSGVALOGSPEIECLPVGALAOMNV  | 1472 |
| Qy | 1547 | SAPTCVPPCGGNLTERRGTTILSPGFBEPYLNLSNCKWIKIIVPBGAGIOIOVSPFTEBOM | 1606 |
| Db | 1473 | SAPTCVPPCGGNLTERRGTTILSPGFBEPYLNLSNCKWIKIIVPBGAGIOIOVSPFTEBOM | 1532 |
| Qy | 1607 | DSLEVPDADNTVYMLGSFGGTTVPALLNNTSNQVLYLHFYSDISVSNAAGFHEKTVGLS   | 1666 |
| Db | 1533 | DSLEVPDADNTVYMLGSFGGTTVPALLNNTSNQVLYLHFYSDISVSNAAGFHEKTVGLS   | 1592 |
| Qy | 1667 | SCPEPAVPSNGVKTGERVLYNDVVSFOCEPGVALQGHAIISCMPTVRNNYPPPLCIAQ    | 1726 |
| Db | 1593 | SCPEPAVPSNGVKTGERVLYNDVVSFOCEPGVALQGHAIISCMPTVRNNYPPPLCIAQ    | 1652 |
| Qy | 1727 | CGGTVEEMEGVILSPGPRGYPVPSNODSMKIALPVGGAHIOPLNFTSEBNHYIEIRNG    | 1786 |
| Db | 1653 | CGGTVEEMEGVILSPGPRGYPVPSNODSMKIALPVGGAHIOPLNFTSEBNHYIEIRNG    | 1712 |
| Qy | 1787 | PYETSRMAGRFSGSELPSLSLSTSHETTVYFHSDSQNRPGFKLEYQAYELOCEPDDEPF   | 1846 |
| Db | 1713 | PYETSRMAGRFSGSELPSLSLSTSHETTVYFHSDSQNRPGFKLEYQAYELOCEPDDEPF   | 1772 |
| Qy | 1847 | ANGIVRGAGYNVQSVTFECLPGYOLTGHPVLTTCQHTNNDHPLPKCEVPCCGNTSS      | 1906 |
| Db | 1773 | ANGIVRGAGYNVQSVTFECLPGYOLTGHPVLTTCQHTNNDHPLPKCEVPCCGNTSS      | 1832 |
| Qy | 1907 | NGTVYSPGFPSPYSSODCVMWLTVPICGHGVRLNLSLQTEBPGDFTTIDGPOQVAPRL    | 1966 |
| Db | 1833 | NGTVYSPGFPSPYSSODCVMWLTVPICGHGVRLNLSLQTEBPGDFTTIDGPOQVAPRL    | 1892 |
| Qy | 1967 | GVFTSRMAKKTVOSSSNQVLLKFRDAATGAI FALAFSAPLTKCPPTILPNAEVTEN     | 2026 |
| Db | 1893 | GVFTSRMAKKTVOSSSNQVLLKFRDAATGAI FALAFSAPLTKCPPTILPNAEVTEN     | 1952 |
| Qy | 2027 | EEFNIGDIVRYCLPGFTLVGNELTCKLGTVYQFEGBPPICEVHCETNELTIDSTGVL     | 2086 |
| Db | 1953 | EEFNIGDIVRYCLPGFTLVGNELTCKLGTVYQFEGBPPICEVHCETNELTIDSTGVL     | 2012 |
| Qy | 2087 | SQSYRGSPORFQCSMIVRYEPDYNISLTJTEYFLEKQVDEFEIPDGPSSQSLPKXLSG    | 2146 |
| Db | 2013 | SQSYRGSPORFQCSMIVRYEPDYNISLTJTEYFLEKQVDEFEIPDGPSSQSLPKXLSG    | 2072 |
| Qy | 2147 | NYSAPLIVTSSNSVYLRMSDHAYNRKGEKIRASAPCSLPRAPLHGLTIGOTSTPGG      | 2206 |
| Db | 2073 | NYSAPLIVTSSNSVYLRMSDHAYNRKGEKIRASAPCSLPRAPLHGLTIGOTSTPGG      | 2132 |
| Qy | 2207 | SIHFQCNAGYRLVGHSMALCTRRHPQGYHLSSEALPLCOALSGLPEAPKXGMVFGKEYTV  | 2266 |
| Db | 2133 | SIHFQCNAGYRLVGHSMALCTRRHPQGYHLSSEALPLCOALSGLPEAPKXGMVFGKEYTV  | 2192 |
| Qy | 2267 | GTAXVYSCSGYHLQAGAEATACLDITGLMSNNRVPOQCVNYPCPDVSSISVEHGRRLI    | 2326 |
| Db | 2193 | GTAXVYSCSGYHLQAGAEATACLDITGLMSNNRVPOQCVNYPCPDVSSISVEHGRRLI    | 2252 |
| Qy | 2327 | FETOYQFOQMLICDPGGYVYTGQRYVIRCOANGKMSLGDSTPCRIISCEGLPIPPGHR    | 2386 |
| Db | 2253 | FETOYQFOQMLICDPGGYVYTGQRYVIRCOANGKMSLGDSTPCRIISCEGLPIPPGHR    | 2312 |
| Qy | 2387 | IGTLVYGAATAFSCNSGYTLVGSRYRECMANGLMSGSBYRCLAGHGTDEPIVNGHNG     | 2446 |
| Db | 2313 | IGTLVYGAATAFSCNSGYTLVGSRYRECMANGLMSGSBYRCLAGHGTDEPIVNGHNG     | 2372 |
| Qy | 2447 | ENYSYRGSVYVYQCNAGRLLIGMSYRICOQDHMSGKTPFCVPIITCGHGRPVNGLTQGNQ  | 2506 |
| Db | 2373 | ENYSYRGSVYVYQCNAGRLLIGMSYRICOQDHMSGKTPFCVPIITCGHGRPVNGLTQGNQ  | 2413 |
| Qy | 2507 | FNLNDVVKFVNCNPGYNAEGAARSQLASQMSDMLPTCRIINCTDPGHQENSRQVYASG    | 2566 |
| Db | 2414 | FNLNDVVKFVNCNPGYNAEGAARSQLASQMSDMLPTCRIINCTDPGHQENSRQVYASG    | 2413 |
| Qy | 2567 | PHRFSGTTVSYRNCNHFYLLGTPVLSQGDGTWDRBRPOCLLVSCGHPSPPHSONSGD     | 2626 |
| Db | 2414 | PHRFSGTTVSYRNCNHFYLLGTPVLSQGDGTWDRBRPOCLLVSCGHPSPPHSONSGD     | 2431 |

QY 2627 SYTVGAIVVYSCIGKRTLVGNSTRMCGLDGHWGSLPHCSGTSVGVCGDPGPIPAHGIRLG 2686  
 DB 2432 SYTVGAIVVYSCIGKRTLVGNSTRMCGLDGHWGSLPHCSGTSVGVCGDPGPIPAHGIRLG 2491  
 QY 2687 DSFPDPTWRFSCGAGHVLRGSSERTCOANGSWSSGQPECGVISCANPQTPSNARVPSD 2746  
 DB 2492 DSFPDPTWRFSCGAGHVLRGSSERTCOANGSWSSGQPECGVISCANPQTPSNARVPSD 2551  
 QY 2747 GLVFSSTIYECREGVATGTLRSRCHSVNVTGSDPCLVINCDDPCLPANGRLGNDP 2806  
 DB 2552 GLVFSSTIYECREGVATGTLRSRCHSVNVTGSDPCLVINCDDPCLPANGRLGNDP 2611  
 QY 2807 RNKTVTYTQCVGYMMESHRSVLSCTDRTNWNGTKPVCKALMCKPPLINGKVGSDP 2866  
 DB 2612 RNKTVTYTQCVGYMMESHRSVLSCTDRTNWNGTKPVCKALMCKPPLINGKVGSDP 2671  
 QY 2867 MNGSSVTVACLEGYOLSLPAYTCGNGSNMCELPOCPVPCGDPGVRSRRRDRGFSY 2926  
 DB 2672 MNGSSVTVACLEGYOLSLPAYTCGNGSNMCELPOCPVPCGDPGVRSRRRDRGFSY 2731  
 QY 2927 RSSVSFSCHPPLVLVGSRRRCQSDGTWGTQPSCTIDPTLTTCADPVPQPGIOMNSQY 2986  
 DB 2732 RSSVSFSCHPPLVLVGSRRRCQSDGTWGTQPSCTIDPTLTTCADPVPQPGIOMNSQY 2791  
 QY 2987 QVGSVTVLRCCQKGYLLQGSTTRTCLPNTWGTGTPPDVPHCRQPTPTANVAGALDPS 3046  
 DB 2792 QVGSVTVLRCCQKGYLLQGSTTRTCLPNTWGTGTPPDVPHCRQPTPTANVAGALDPS 2851  
 QY 3047 MGYTLI 3052  
 DB 2852 MGYTLI 2857

RESULT 8  
 ID ABG79169 standard; protein; 2669 AA.  
 AC ABG79169;  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Human cub and sushi domain containing protein #2.  
 XX  
 KW Human; NOX; neurodegenerative disease; Alzheimer's disease; anxiety;  
 KW Parkinson's disease; Huntington's disease; neurological disorder;  
 KW schizophrenia; manic depression; mental retardation; angina pectoris;  
 KW cardiovascular disease; acute heart failure; myocardial infarction;  
 KW muscular disease; muscular disorder; retinal disease; photoreception;  
 KW deafness; keratinisation disorder; cancer; ovarian cancer; melanoma;  
 KW immunological disorder; inflammatory disease; immune disease; diabetes;  
 KW bacterial infection; fungal infection; protozoal infection; obesity;  
 KW viral infection; reproductive system disorder; metabolic disturbance;  
 KW anorexia; wasting disorder; chronic disease; infectious disease;  
 KW dyslipidaemia; cub; sushi; myelin; von Willebrand factor; kielin;  
 KW semaphorin; serine/threonine protein kinase; TGF-beta binding;  
 KW mas proto-oncogene; ribonuclease pancreatic precursor; aminotransferase;  
 KW colloid-like 2; cysteine sulfenic acid decarboxylase.  
 OS Homo sapiens.  
 XX  
 PN WO200264791-A2.  
 XX  
 PD 22-AUG-2002.  
 XX  
 PF 10-DEC-2001; 2001MO-US048369.  
 XX  
 PR 08-DEC-2000; 2000US-0254329P.  
 PR 14-DEC-2000; 2000US-0255648P.  
 PR 15-MAY-2001; 2001US-0291037P.  
 PR 08-JUN-2001; 2001US-0297173P.  
 PR 08-JUN-2001; 2001US-0309258P.  
 PR 29-AUG-2001; 2001US-0315639P.  
 PR 01-OCT-2001; 2001US-0326393P.

XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Alsebrook JP, Anderson DM, Burgess CE, Boldog FI, Casman SJ,  
 PI Colman SP, Edinger SR, Ellerman K, Gerlach V, Gorman L, Grosse WM,  
 PI Guo X, Herrmann JJ, Kekuda R, Lepley DM, Li L, Macdougall JR,  
 PI Millet I, Pena CE, Peyman JA, Rastelli L, Rieger DK, Shinkets RA,  
 PI Sathusen G, Splet KA, Stone DJ, Tchernev VT, Vernet CAM, Voss EZ,  
 PI Zernhusen BD, Zhong H, Zhong M,  
 XX  
 DR WPI; 2002-643486/69.  
 DR N-PDB; AB564376.  
 XX  
 PT New NOX polypeptides and polynucleotides useful for treating or  
 PT preventing e.g. neurodegenerative diseases, neurological disorders,  
 PT cardiovascular diseases, muscular diseases and disorders, or  
 PT immunological diseases.  
 XX  
 PS Claim 1; Page 16-17; 299pp; English.

CC The present invention relates to new NOX polypeptides. The polypeptides,  
 CC polynucleotides and antibodies are useful in the manufacture of a  
 CC medicament for treating or preventing neurodegenerative diseases (e.g.  
 CC Alzheimer's disease, Parkinson's disease, or Huntington's disease),  
 CC neurological disorders (e.g. anxiety, schizophrenia, manic depression or  
 CC mental retardation), cardiovascular disease (e.g. acute heart failure, or  
 CC angina pectoris or myocardial infarction), muscular diseases and  
 CC disorders, retinal diseases (including those involving photoreception,  
 CC deafness and keratinisation disorders), cancer (e.g. ovarian cancer or  
 CC melanoma), immunological disorders, inflammatory and immune diseases,  
 CC bacterial, fungal, protozoal and viral infections, and reproductive  
 CC system disorders. The proteins of the invention may be used to screen  
 CC drugs or compounds that modulate the NOX protein activity or expression,  
 CC as well as to treat disorders characterised by insufficient or excessive  
 CC production of NOX protein or protein forms that have decreased or  
 CC aberrant activity compared to NOX wild type protein, such as diabetes,  
 CC obesity, metabolic disturbances associated with obesity, anorexia and  
 CC wasting disorders associated with chronic diseases and various cancers,  
 CC infectious diseases and various dyslipidaemias. The nucleic acid  
 CC sequences of the invention may be used in chromosome mapping, identifying  
 CC an individual from minute biological samples (tissue typing), and in  
 CC forensic identification of a biological sample. The present amino acid  
 CC sequence represents a NOX protein of the invention  
 XX  
 SQ Sequence 2669 AA;

Query Match 83.3%; Score 14142.5; DB 5; Length 2669;  
 Best Local Similarity 97.7%; Pred. No. 0;  
 Matches 2605; Conservative 1; Mismatches 2; Indels 59; Gaps 3;

QY 1 MAGAPPALLLPCLISDCCASNORHSGVGPSELYVKQIELKSRGVKMPKXDNQKTS 60  
 DB 1 MAGAPPALLLPCLISDCCASNORHSGVGPSELYVKQIELKSRGVKMPKXDNQKTS 60  
 QY 61 VLTQGVSGQHNNCPDPCIGPERGRKLGSDPRLGSSVQFTNBEVDLQGRITCMKXSDM 120  
 DB 61 VLTQGVSGQHNNCPDPCIGPERGRKLGSDPRLGSSVQFTNBEVDLQGRITCMKXSDM 120  
 QY 121 FAASDHRPVCRARMCDAAHRRGSGITTSNPFIOYDNNAHCWITTLNPSVYIKLAF 180  
 DB 121 FAASDHRPVCRARMCDAAHRRGSGITTSNPFIOYDNNAHCWITTLNPSVYIKLAF 180  
 QY 181 EFDLBERGYDTLVGGDGDQDKTVLYMSQNASDSDPHTFGSRIPEMSGDDIMROKMYL 240  
 DB 181 EFDLBERGYDTLVGGDGDQDKTVLYMSQNASDSDPHTFGSRIPEMSGDDIMROKMYL 240  
 QY 241 EICRDISSDARSGSVRKSPKTSNAVLYAPGTETIQSGCGDGPDPAYRRRSGRPHHD 300  
 DB 241 EICRDISSDARSGSVRKSPKTSNAVLYAPGTETIQSGCGDGPDPAYRRRSGRPHHD 300  
 QY 301 TLKFECPAFELVGVQALITCKKNNQMSAKKPGVSCFNFTSPSGVULSPVYPPEDYGNH 360  
 DB 301 TLKFECPAFELVGVQALITCKKNNQMSAKKPGVSCFNFTSPSGVULSPVYPPEDYGNH 360

QY 361 LHCWMLIARPEBRIHLAFNDIDVEPODFLVIKOGATAEAPVLGTFESGNOLPSSITSSG 420  
 Db 361 LHCWMLIARPEBRIHLAFNDIDVEPODFLVIKOGATAEAPVLGTFESGNOLPSSITSSG 420  
 QY 421 HVARLEFQDHSSTGKRGFNITFTFRNECEPCDPGVNPKRGFGDSLQJGSSISPLCDEGF 480  
 Db 421 HVARLEFQDHSSTGKRGFNITFTFRNECEPCDPGVNPKRGFGDSLQJGSSISPLCDEGF 480  
 QY 481 LGTQGSSTITCVLKESGVWNSAVLRCAPOCGHLTSPSGTILSPGHPGTYKALSCAMV 540  
 Db 481 LGTQGSSTITCVLKESGVWNSAVLRCAPOCGHLTSPSGTILSPGHPGTYKALSCAMV 540  
 QY 541 IEAOPGPIKTPRFRFTEVNDTLAVRDGRTYSAPLIGVYHGVOQOPLISTSNLYL 600  
 Db 541 IEAOPGPIKTPRFRFTEVNDTLAVRDGRTYSAPLIGVYHGVOQOPLISTSNLYL 600  
 QY 601 FSTDKSHSDIGFOLRYETITLQSDHCLDPCI.PVNGORHNDPVYALVTSCDSGYTLSD 660  
 Db 601 FSTDKSHSDIGFOLRYETITLQSDHCLDPCI.PVNGORHNDPVYALVTSCDSGYTLSD 660  
 QY 661 GEPLECEPNFQWSPALPSCALCCGFIQSSGTTLSFGPDPFYNNINCTWIIETSHGK 720  
 Db 661 GEPLECEPNFQWSPALPSCALCCGFIQSSGTTLSFGPDPFYNNINCTWIIETSHGK 720  
 QY 721 VFFPFTFHLESBGHLYLITENGSFTQPLRQLTGSRL.PAPISAGLYGNFTAOVRFTSDF 780  
 Db 721 VFFPFTFHLESBGHLYLITENGSFTQPLRQLTGSRL.PAPISAGLYGNFTAOVRFTSDF 780  
 QY 781 MSYGFNITSEYDLFCEBEPEVAYSIRKGLQGVDDTLTFCSPGTYLEGTAIRITCLG 840  
 Db 781 MSYGFNITSEYDLFCEBEPEVAYSIRKGLQGVDDTLTFCSPGTYLEGTAIRITCLG 840  
 QY 841 GRRRLMSSPLPRCAECGNSVTGTQGTLSLSPFNNTNNHECTIYSIQTOPGKIQOLKAR 900  
 Db 841 GRRRLMSSPLPRCAECGNSVTGTQGTLSLSPFNNTNNHECTIYSIQTOPGKIQOLKAR 900  
 QY 901 AFELSEBDVLKVYDGNNSARLLGVFSHSEMGVTLTSSSLMDITTDENSTSKPEL 960  
 Db 901 AFELSEBDVLKVYDGNNSARLLGVFSHSEMGVTLTSSSLMDITTDENSTSKPEL 960  
 QY 961 HFSFELIKCEDPGTPKFGYKVBDEGHFAGSSVSFCDPGYSLAGESELLCLSGERTW 1020  
 Db 961 HFSFELIKCEDPGTPKFGYKVBDEGHFAGSSVSFCDPGYSLAGESELLCLSGERTW 1020  
 QY 1021 RPLPTCYAECGGTYRGEVSGVLSPGYPARYENHLNCITWIEBAGCTTGLHFLVPTDEE 1080  
 Db 1021 RPLPTCYAECGGTYRGEVSGVLSPGYPARYENHLNCITWIEBAGCTTGLHFLVPTDEE 1080  
 QY 1081 VHDVLRITWDGVEBEGVLLKELSGPALPKDLHSTFNSVULOFTSTKOGFALQFVSST 1140  
 Db 1081 VHDVLRITWDGVEBEGVLLKELSGPALPKDLHSTFNSVULOFTSTKOGFALQFVSST 1140  
 QY 1141 ATSCNDGIPONGSRSDSMEAGDSVTFQCDPGYALOGSABEISCVKLENFEPFOPSPPTC 1200  
 Db 1141 ATSCNDGIPONGSRSDSMEAGDSVTFQCDPGYALOGSABEISCVKLENFEPFOPSPPTC 1200  
 QY 1201 IAPCGDLTGPSSGVILSPNYPPEYP.PPGKECDMKVTASPDVIALVFNIPMLEPGYDLHI 1260  
 Db 1201 IAPCGDLTGPSSGVILSPNYPPEYP.PPGKECDMKVTASPDVIALVFNIPMLEPGYDLHI 1260  
 QY 1261 YDGDSDLSPLIGSYGQOLPGRITSSNSLFLAFRSDASVSNAPFVIDYENPRESCEFD 1320  
 Db 1261 YDGDSDLSPLIGSYGQOLPGRITSSNSLFLAFRSDASVSNAPFVIDYENPRESCEFD 1320  
 QY 1321 GSINRGTRVSGDLKLGSSVTTYCHGGEVEGTSTLSCILGPDGPRVWNNPRPVCTACGG 1380  
 Db 1321 GSINRGTRVSGDLKLGSSVTTYCHGGEVEGTSTLSCILGPDGPRVWNNPRPVCTACGG 1380  
 QY 1381 QYVSGDGVLSPNYPONTSGOICLYFVTVBKDYVFGQAF.FHTALNDVVEVHDSQH 1440  
 Db 1381 QYVSGDGVLSPNYPONTSGOICLYFVTVBKDYVFGQAF.FHTALNDVVEVHDSQH 1440

QY 1441 SRLISSLSGHTGESLPLATSNQVLIFKSAKGLAPARGFFHYQAVPRTSATOCSSVPER 1500  
 Db 1441 SRLISSLSGHTGESLPLATSNQVLIFKSAKGLAPARGFFHYQAVPRTSATOCSSVPER 1500  
 QY 1501 RYKRLGSDFSVGAIVRECNISGVALQGSBEIECLVPGLAOMVSAPTCVPCGANT 1560  
 Db 1501 RYKRLGSDFSVGAIVRECNISGVALQGSBEIECLVPGLAOMVSAPTCVPCGANT 1560  
 QY 1561 ERGGTILSPGPEPYLNSLNCWKIIVREAGQIOIVNSVTEOMNDSLEVPDADTVT 1620  
 Db 1561 ERGGTILSPGPEPYLNSLNCWKIIVREAGQIOIVNSVTEOMNDSLEVPDADTVT 1620  
 QY 1621 MLGSFGTVPALLNSTSNQLYLHFSYSDISVSAAGFHEKTYGLSSCEPAPVSNVKT 1680  
 Db 1621 MLGSFGTVPALLNSTSNQLYLHFSYSDISVSAAGFHEKTYGLSSCEPAPVSNVKT 1680  
 QY 1681 GERLYVNDVVSFOCEPQYALQGHANISCMGTYVRMNYPPPLCTAOCGGTVEEMGVILS 1740  
 Db 1681 GERLYVNDVVSFOCEPQYALQGHANISCMGTYVRMNYPPPLCTAOCGGTVEEMGVILS 1740  
 QY 1741 PGFPNYPNSMDCSWKIALPVGFGAHQFINSTEPNHDYIELRNGYETSRMNGRPSGS 1800  
 Db 1741 PGFPNYPNSMDCSWKIALPVGFGAHQFINSTEPNHDYIELRNGYETSRMNGRPSGS 1800  
 QY 1801 ELPSLLSTSHETTVFPHSDSONRPGFKLEYAYELOECBPDEPFANGIVRAGYVWQ 1860  
 Db 1801 ELPSLLSTSHETTVFPHSDSONRPGFKLEYAYELOECBPDEPFANGIVRAGYVWQ 1860  
 QY 1861 SVTFECLPGYQLTGHPLVLTQHGSTRNWDHPLKCEVPCGNTTSSNGTVSFGPSPYS 1920  
 Db 1861 SVTFECLPGYQLTGHPLVLTQHGSTRNWDHPLKCEVPCGNTTSSNGTVSFGPSPYS 1920  
 QY 1921 SSQDCWMLIVPFGHGRMLSLLOTPESGSDFTTMDPGQTPRGLCVFRSMAKKTVO 1980  
 Db 1921 SSQDCWMLIVPFGHGRMLSLLOTPESGSDFTTMDPGQTPRGLCVFRSMAKKTVO 1980  
 QY 1981 SSNOVLKFRDAAATGIFALFAFSAVPLTCCPPTILPNAEVTENEEFNIGDIVRCL 2040  
 Db 1981 SSNOVLKFRDAAATGIFALFAFSAVPLTCCPPTILPNAEVTENEEFNIGDIVRCL 2040  
 QY 2041 PGFTLVGNELITCKLGTLYQFEGPPICEVHCPTEBLTDTSTVILISQSYGSGYPOFOTC 2100  
 Db 2041 PGFTLVGNELITCKLGTLYQFEGPPICEVHCPTEBLTDTSTVILISQSYGSGYPOFOTC 2100  
 QY 2101 SWLVRVDPDYNISLTYEYFLSEKQYDEFEI.FDPSGQSLKXLSGNYAPLVITSSNS 2160  
 Db 2101 SWLVRVDPDYNISLTYEYFLSEKQYDEFEI.FDPSGQSLKXLSGNYAPLVITSSNS 2160  
 QY 2161 VYLRWSDHAYNRKGFIRYSAPYCSLPRALHGFILGOTSTPGGSIFHGCNAGYLVG 2220  
 Db 2161 VYLRWSDHAYNRKGFIRYSAPYCSLPRALHGFILGOTSTPGGSIFHGCNAGYLVG 2220  
 QY 2221 HSNALCTRHPOGYHLSSEALPLCOALSCLGPEAPKNGMVGKEYVTGTRAVYSCSEGYL 2280  
 Db 2221 HSNALCTRHPOGYHLSSEALPLCOALSCLGPEAPKNGMVGKEYVTGTRAVYSCSEGYL 2280  
 QY 2281 QAGAEATACLDLDTGLMSNRVPPQCVRESGNGSGSVTCDDVASSIEHGRMLIFETQY 2340  
 Db 2281 QAGAEATACLDLDTGLMSNRVPPQCVRESGNGSGSVTCDDVASSIEHGRMLIFETQY 2340  
 QY 2332 QFOAOLMLICDPPGYTTCGVIRICQANGKMSLDSPTCRHISGGEPLIPNHRIGTILS 2391  
 Db 2332 QFOAOLMLICDPPGYTTCGVIRICQANGKMSLDSPTCRHISGGEPLIPNHRIGTILS 2391  
 QY 2341 QFOAOLMLICDPPGYTTCGVIRICQANGKMSLDSPTCRHISGGEPLIPNHRIGTILS 2400  
 Db 2341 QFOAOLMLICDPPGYTTCGVIRICQANGKMSLDSPTCRHISGGEPLIPNHRIGTILS 2400  
 QY 2392 VYGATAIFSCNSGYTLVGRVRECMANGLWSGSEVCL 2429  
 Db 2392 VYGATAIFSCNSGYTLVGRVRECMANGLWSGSEVCL 2429  
 QY 2401 VYGATAIFSCNSGYTLVGRVRECMANGLWSGSEVCLATQTKLHSIFYKLLFDVLSPPS 2460  
 Db 2401 VYGATAIFSCNSGYTLVGRVRECMANGLWSGSEVCLATQTKLHSIFYKLLFDVLSPPS 2460  
 QY 2430 ---AGHCGETEPLVNGHINENSYRSGSVYVQCNAGPRLIGMSVRIQODHHSKGTFFC 2486  
 Db 2430 ---AGHCGETEPLVNGHINENSYRSGSVYVQCNAGPRLIGMSVRIQODHHSKGTFFC 2486  
 QY 2461 LTRAGHGTETPLVNGHINENSYRSGSVYVQCNAGPRLIGMSVRIQODHHSKGTFFC 2520  
 Db 2461 LTRAGHGTETPLVNGHINENSYRSGSVYVQCNAGPRLIGMSVRIQODHHSKGTFFC 2520  
 QY 2487 V-----PITCGHGNPNVNGLTQCNQFNLDVVFVFCNPGY 2521

```

Db      2521 VHVKKOQLLLLLLLCDDEDDDDSGALTTCGPGNPNVGLQGNQFNLDVVKVVCNPGY 2560
Qy      2522 MAEGARSOCLASGQSMPLPTCRILINCTDGHQNSRQVHASGPHRFSGTTVSYRCN 2561
Db      2561 MAEGARSOCLASGQSMPLPTCRILINCTDGHQNSRQVHASGPHRFSGTTVSYRCN 2640
Qy      2582 HGFYLLGTPVLSCQSGDGTMDRPRPQCL 2608
Db      2641 HGFYLLGTPVLSCQSGDGTMDRPRPQCL 2667

RESULT 9
ADH71140
ID      ADH71140 standard; protein; 2669 AA.
XX
AC      ADH71140;
XX
DT      25-MAR-2004 (first entry)
XX
DE      Human protein of the invention NOV4c SEQ ID NO:36.
XX
KW      human; cytosolic; immunomodulator; neuroprotective; nootropic;
KW      anorectic; antidiabetic; antimicrobial; antihypaemic; gene therapy;
KW      vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
KW      obesity; diabetes; infectious disease; metabolic syndrome X;
KW      dyslipidaemia.
XX
OS      Homo sapiens.
XX
PN      WO000102155-A2.
XX
PD      11-DEC-2003.
XX
PF      03-JUN-2003; 2003WO-US017430.
XX
PR      03-JUN-2002; 2002US-0385120P.
PR      04-JUN-2002; 2002US-0385784P.
PR      05-JUN-2002; 2002US-0386041P.
PR      05-JUN-2002; 2002US-0386047P.
PR      06-JUN-2002; 2002US-0386376P.
PR      06-JUN-2002; 2002US-0386453P.
PR      06-JUN-2002; 2002US-0386644P.
PR      06-JUN-2002; 2002US-0387016P.
PR      07-JUN-2002; 2002US-0386796P.
PR      07-JUN-2002; 2002US-0386816P.
PR      07-JUN-2002; 2002US-0386931P.
PR      07-JUN-2002; 2002US-0386942P.
PR      07-JUN-2002; 2002US-0386971P.
PR      08-JUN-2002; 2002US-0387262P.
PR      08-JUN-2002; 2002US-0296360P.
PR      10-JUN-2002; 2002US-0387400P.
PR      10-JUN-2002; 2002US-0387535P.
PR      11-JUN-2002; 2002US-0387610P.
PR      11-JUN-2002; 2002US-0387625P.
PR      11-JUN-2002; 2002US-0387634P.
PR      11-JUN-2002; 2002US-0387668P.
PR      11-JUN-2002; 2002US-0387696P.
PR      11-JUN-2002; 2002US-0387702P.
PR      11-JUN-2002; 2002US-0387836P.
PR      11-JUN-2002; 2002US-0387859P.
PR      12-JUN-2002; 2002US-0387933P.
PR      12-JUN-2002; 2002US-0387934P.
PR      12-JUN-2002; 2002US-0387960P.
PR      12-JUN-2002; 2002US-0388022P.
PR      12-JUN-2002; 2002US-0388096P.
PR      13-JUN-2002; 2002US-0389123P.
PR      14-JUN-2002; 2002US-0389118P.
PR      14-JUN-2002; 2002US-0389120P.
PR      14-JUN-2002; 2002US-0389144P.
PR      14-JUN-2002; 2002US-0389146P.
PR      17-JUN-2002; 2002US-0389129P.
PR      17-JUN-2002; 2002US-0389142P.

```

```

PR      18-JUN-2002; 2002US-0389844P.
PR      19-JUN-2002; 2002US-0390066P.
PR      19-JUN-2002; 2002US-0390209P.
PR      21-JUN-2002; 2002US-0390763P.
PR      17-JUL-2002; 2002US-0396706P.
PR      06-AUG-2002; 2002US-0401628P.
PR      09-AUG-2002; 2002US-0402156P.
PR      09-AUG-2002; 2002US-0402256P.
PR      09-AUG-2002; 2002US-0402389P.
PR      12-AUG-2002; 2002US-0402786P.
PR      12-AUG-2002; 2002US-0402816P.
PR      12-AUG-2002; 2002US-0402821P.
PR      12-AUG-2002; 2002US-0402832P.
PR      13-AUG-2002; 2002US-0403448P.
PR      13-AUG-2002; 2002US-0403459P.
PR      13-AUG-2002; 2002US-0403531P.
PR      13-AUG-2002; 2002US-0403532P.
PR      13-AUG-2002; 2002US-0403563P.
PR      13-AUG-2002; 2002US-0406317P.
PR      15-AUG-2002; 2002US-0406317P.
PR      15-AUG-2002; 2002US-0406182P.
PR      26-AUG-2002; 2002US-0406182P.
PR      26-AUG-2002; 2002US-0406355P.
PR      27-AUG-2002; 2002US-0406240P.
PR      12-SEP-2002; 2002US-0410084P.
PR      20-SEP-2002; 2002US-0412528P.
PR      23-SEP-2002; 2002US-0412731P.
PR      30-SEP-2002; 2002US-0414801P.
PR      30-SEP-2002; 2002US-0414839P.
PR      30-SEP-2002; 2002US-0414840P.
PR      30-SEP-2002; 2002US-0414954P.
PR      09-OCT-2002; 2002US-0417186P.
PR      09-OCT-2002; 2002US-0417406P.
PR      23-OCT-2002; 2002US-0420639P.
PR      28-OCT-2002; 2002US-0421156P.
PR      31-OCT-2002; 2002US-0422690P.
PR      01-NOV-2002; 2002US-0423130P.
PR      05-NOV-2002; 2002US-0423798P.
PR      05-NOV-2002; 2002US-0423798P.
PR      12-NOV-2002; 2002US-0425453P.

(PURA) CYPAGEN CORP.
XX
XX      Alsobrook JP, Alvarez E, Anderson DW, Boldog FL, Casman SJ;
PI      Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
PI      Ettenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
PI      Guev VY, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, Macdougall JR;
PI      MacLachlan T, Malyankar UM, Mezick AJ, Millet I, Mishra VS;
PI      Padigaru W, Paturajan M, Pena CE, Peyman JA, Raha D, Raetelli L;
PI      Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkets RA;
PI      Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;
XX      Zhong H;
XX
XX      WPI; 2004-081935/08.
DR      N-PSDB; ADH71139.
XX
XX      New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT      treating NOVX-associated disorders, e.g. cancer, diabetes, infection or
PT      obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
XX
XX      Example 4; SEQ ID NO 36; 1880bp; English.
XX
XX      The invention relates to a novel isolated polypeptide (NOVX). A
CC      polypeptide of the invention has cytosratic, immunomodulator,
CC      neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
CC      antihypaemic activity, and may have a use in gene therapy, and as a
CC      vaccine. The polypeptides are encoded by NOVX polynucleotides comprising
CC      any of the 303 fully defined nucleotide sequences given in the
CC      specification. The polypeptide is useful in the manufacture of a
CC      medicament for treating a syndrome associated with a human disease. The
CC      polypeptide, polynucleotide and antibody are useful in diagnosing,
CC      treating or preventing NOVX-associated disorders, e.g. cancer, cachexia,
CC      Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
CC      diseases, metabolic syndrome X or dyslipidaemia. The nucleic acids are

```

CC further used as hybridisation probes, in chromosome mapping, tissue  
 CC typing, preventive medicine, and pharmacogenetics. The present sequence  
 CC represents a NOXK polypeptide of the invention.

XX Sequence 2669 AA;

Query Match 83.3%; Score 14142.5; DB 8; Length 2669;  
 Best Local Similarity 97.7%; Pred. No. 0;  
 Matches 2605; Conservative 1; Mismatches 2; Indels 59; Gaps 3;

QY 1 MAGAPPALLPCGLISDCCASQNRHSVGVSELVKKQIELKSGVKKMPSKONSQKTS 60  
 DB 1 MAGAPPALLPCGLISDCCASQNRHSVGVSELVKKQIELKSGVKKMPSKONSQKTS 60  
 QY 61 VLTQVGVSGHNMCPDGGIPEGRKLGSDPRLGSSVQFTCNEGVDLOGSKRITCMKSDM 120  
 DB 61 VLTQVGVSGHNMCPDGGIPEGRKLGSDPRLGSSVQFTCNEGVDLOGSKRITCMKSDM 120  
 QY 121 FAAMSDRRPVCRAAMCAHLRGPSCIITSPNFIQYNNNAHCWIIITALNPSKVIKLAPE 180  
 DB 121 FAAMSDRRPVCRAAMCAHLRGPSCIITSPNFIQYNNNAHCWIIITALNPSKVIKLAPE 180  
 QY 181 EFDLERYDTLTVGSGDQDQKTVLYMSQNAQSDSPHTPGSRIPESMGDIWRQKXTVL 240  
 DB 181 EFDLERYDTLTVGSGDQDQKTVLYMSQNAQSDSPHTPGSRIPESMGDIWRQKXTVL 240  
 QY 241 EICRDISSDARSVSKSPKTSNAVELVAPGTEIBGSGCDPDIPIAYGRREGRFHGD 300  
 DB 241 EICRDISSDARSVSKSPKTSNAVELVAPGTEIBGSGCDPDIPIAYGRREGRFHGD 300  
 QY 301 TLKFCOPAFELVQKAITCOKNQMSAKKPGCVFSCFNPFTSPGVLSPNYPEDYGNH 360  
 DB 301 TLKFCOPAFELVQKAITCOKNQMSAKKPGCVFSCFNPFTSPGVLSPNYPEDYGNH 360  
 QY 361 LHCWMLIABPESRIHLAFNDIYVEPODFLVIKQATAPAVLGTSGNQLPSSITSG 420  
 DB 361 LHCWMLIABPESRIHLAFNDIYVEPODFLVIKQATAPAVLGTSGNQLPSSITSG 420  
 QY 421 HVARLEQTHSTKRGFNITFTFRNECPDPVAVNGKRFQDLSLQSSISLTCBEGF 480  
 DB 421 HVARLEQTHSTKRGFNITFTFRNECPDPVAVNGKRFQDLSLQSSISLTCBEGF 480  
 QY 481 LGTQSGEITTCVLEKGSVWNSAVLRCBAPCGHLTSPSGTILSPGMGFKDALSCAW 540  
 DB 481 LGTQSGEITTCVLEKGSVWNSAVLRCBAPCGHLTSPSGTILSPGMGFKDALSCAW 540  
 QY 541 IEAOPGYPIKITPDRFTEVNYDTLAVRDGRVSAPLIGVYHGTQVQPLISTSNLYL 600  
 DB 541 IEAOPGYPIKITPDRFTEVNYDTLAVRDGRVSAPLIGVYHGTQVQPLISTSNLYL 600  
 QY 601 FSTDKSHSDIGFQURYTEITLQSDHCLDPGI PVNGQHRGNDPVYGAIVTFSCSGYTLSD 660  
 DB 601 FSTDKSHSDIGFQURYTEITLQSDHCLDPGI PVNGQHRGNDPVYGAIVTFSCSGYTLSD 660  
 QY 661 GEPLECPNFMQMSALPSCALCGGFIQSSGTTILSGPPDFYNNLNCMIITETSHGK 720  
 DB 661 GEPLECPNFMQMSALPSCALCGGFIQSSGTTILSGPPDFYNNLNCMIITETSHGK 720  
 QY 721 VFTFHTFHLSESHDYLITENGSTFQPLRLQLTGSRPAPISAGLYNFTAAQVRFISDFS 780  
 DB 721 VFTFHTFHLSESHDYLITENGSTFQPLRLQLTGSRPAPISAGLYNFTAAQVRFISDFS 780  
 QY 781 MSYEGFNTTSEYDLEPCBEPEVAYISIRKGLQGVADTILFSCFPGYREGTARITCLG 840  
 DB 781 MSYEGFNTTSEYDLEPCBEPEVAYISIRKGLQGVADTILFSCFPGYREGTARITCLG 840  
 QY 841 GRRRLMSPLPRCAEAGNSVTGQGLTSLSPNPNVNNNHECYSYLQTPGKGIOLKAR 900  
 DB 841 GRRRLMSPLPRCAEAGNSVTGQGLTSLSPNPNVNNNHECYSYLQTPGKGIOLKAR 900  
 QY 901 AFELSEGDVLYKVDGNNSARLLGVFSHSEMMGVTLNSTSSLWLDITDAENTSCKFEL 960  
 DB 901 AFELSEGDVLYKVDGNNSARLLGVFSHSEMMGVTLNSTSSLWLDITDAENTSCKFEL 960

QY 961 HFSSFEIKCEDDGTPEKGYKVHDEGHFAGSSVSFSCDPGYSLRGSEILCLSGERRTWD 1020  
 DB 961 HFSSFEIKCEDDGTPEKGYKVHDEGHFAGSSVSFSCDPGYSLRGSEILCLSGERRTWD 1020  
 QY 1021 RPLPTCAECCGCTVRGEVSGQVLSPGYPAYENLNCIMTIEACGCTIGLHFLVPFTEE 1080  
 DB 1021 RPLPTCAECCGCTVRGEVSGQVLSPGYPAYENLNCIMTIEACGCTIGLHFLVPFTEE 1080  
 QY 1081 VHDVLRITWDGPVSEGVILKELSGPALPKDIHSTFNSVVLQSTDFPFSKOGFAIQESVST 1140  
 DB 1081 VHDVLRITWDGPVSEGVILKELSGPALPKDIHSTFNSVVLQSTDFPFSKOGFAIQESVST 1140  
 QY 1141 ATSCNDPDIPONSRSRSDSWEAGDSTVFOCDPGALQGSALSCVKEKFNFPQSPPTC 1200  
 DB 1141 ATSCNDPDIPONSRSRSDSWEAGDSTVFOCDPGALQGSALSCVKEKFNFPQSPPTC 1200  
 QY 1201 IAPCGDLTGPSCVITLSPNYPPEYPPGKECDMKVTYSPDVIALVFNI FNLBPGYDLHI 1260  
 DB 1201 IAPCGDLTGPSCVITLSPNYPPEYPPGKECDMKVTYSPDVIALVFNI FNLBPGYDLHI 1260  
 QY 1261 YDGRDLSPLIGSFYGSQLPGRIESSNSLFLAFRSDASVSNAGFVIDYENPRESCFDP 1320  
 DB 1261 YDGRDLSPLIGSFYGSQLPGRIESSNSLFLAFRSDASVSNAGFVIDYENPRESCFDP 1320  
 QY 1321 GSTIKNGRVSGLKLGSSVTVYCHGGYEVGCTSLGCLPDDKPVNNNRPVCTAPCGG 1380  
 DB 1321 GSTIKNGRVSGLKLGSSVTVYCHGGYEVGCTSLGCLPDDKPVNNNRPVCTAPCGG 1380  
 QY 1381 QYVGSQGVLSPNVQNYTSGQICLVFVTVPKQYVVGQAFPHTLANDVEVHDSQH 1440  
 DB 1381 QYVGSQGVLSPNVQNYTSGQICLVFVTVPKQYVVGQAFPHTLANDVEVHDSQH 1440  
 QY 1441 SRIILSLSGSHTEBSLPLATSNQVLIKPSAKGLAPARGHFVYQAVPRISATQCSSVPEP 1500  
 DB 1441 SRIILSLSGSHTEBSLPLATSNQVLIKPSAKGLAPARGHFVYQAVPRISATQCSSVPEP 1500  
 QY 1501 RYGRKLGSDPSGAIYRFECSGVALQGSPEICLPVPGALQMNVSAPTCVPPCGNLT 1560  
 DB 1501 RYGRKLGSDPSGAIYRFECSGVALQGSPEICLPVPGALQMNVSAPTCVPPCGNLT 1560  
 QY 1561 ERGCTILSPGFPEPYLNSLNCWKIIVPEBAGIQI QVVSFVTEQNMDSLEVPDQADTVT 1620  
 DB 1561 ERGCTILSPGFPEPYLNSLNCWKIIVPEBAGIQI QVVSFVTEQNMDSLEVPDQADTVT 1620  
 QY 1621 MUGSFSGTVPALLNSTSNQYLHFYSDISVSAAGFHELYKTVGLSSCPBPAVPSNGVT 1680  
 DB 1621 MUGSFSGTVPALLNSTSNQYLHFYSDISVSAAGFHELYKTVGLSSCPBPAVPSNGVT 1680  
 QY 1681 GERLYLVNDVVSFOCEBRYALQGHANHS CMGTYRNMVPPPLCIAQGGTVEEMEGYILS 1740  
 DB 1681 GERLYLVNDVVSFOCEBRYALQGHANHS CMGTYRNMVPPPLCIAQGGTVEEMEGYILS 1740  
 QY 1741 PGFPYVPSNMDCSWKIALPVFGCAHIQFANFSTEPNHDYIIRNGPYETSRMGRSGS 1800  
 DB 1741 PGFPYVPSNMDCSWKIALPVFGCAHIQFANFSTEPNHDYIIRNGPYETSRMGRSGS 1800  
 QY 1801 ELPSLSLSTSHETTVYFHSQNSQNRPGFKLEYAYELQECPRDEPFANGIVRAGYVVGQ 1860  
 DB 1801 ELPSLSLSTSHETTVYFHSQNSQNRPGFKLEYAYELQECPRDEPFANGIVRAGYVVGQ 1860  
 QY 1861 SYTFECLPGYQLGHPPLTQHGTRNWDPLKCEVPCGGNITSSNGTYYSGFPSYPS 1920  
 DB 1861 SYTFECLPGYQLGHPPLTQHGTRNWDPLKCEVPCGGNITSSNGTYYSGFPSYPS 1920  
 QY 1921 SSQDCWLVITVPIGHGVALNLSTLQTEPSGDFITIMDGPQTAARLGVFRSMAKKTQVS 1980  
 DB 1921 SSQDCWLVITVPIGHGVALNLSTLQTEPSGDFITIMDGPQTAARLGVFRSMAKKTQVS 1980  
 QY 1981 SSNQVLLKFRDAATGGAIFAIAFSAYPLTKCPPTLIPNAEVVTEENEFNIGDI VRYRCL 2040  
 DB 1981 SSNQVLLKFRDAATGGAIFAIAFSAYPLTKCPPTLIPNAEVVTEENEFNIGDI VRYRCL 2040



PA (CURA-) CURAGEN CORP.  
 XX  
 PI Alabobcock JR, Alvarez E, Anderson DW, Boldog FI, Caeman SJ;  
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;  
 PI Ettenberg S, Gangolli EA, Gerlach VL, Gorman L, Gunther E, Guo X;  
 PI Gusev VY, Hermann JU, Ji W, Kekuda R, Li L, Liu X, MacDougall JR;  
 PI Maclellan T, Malpanker UM, Mexick AJ, Millet I, Mishra VS;  
 PI Pedigaru M, Patutanjan M, Pena CE, Peyman JA, Raha D, Rastelli L;  
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shimkets RA;  
 PI Slatheon G, Spytek KA, Stone DU, Vernet CM, Voss EZ, Zhong M,  
 PI Zhong H;  
 DR WPI: 2004-081935/08.  
 DR N-PSDB; ADH71141.  
 PT New NOVX polypeptides and nucleic acid molecules useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or  
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.  
 XX  
 PS Example 4; SEQ ID NO 38; 1880bp; English.  
 XX  
 CC The invention relates to a novel isolated polypeptide (NOVX). A  
 CC polypeptide of the invention has cytosolic, immunomodulator,  
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and  
 CC antipneumatic activity, and may have a use in gene therapy, and as a  
 CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising  
 CC any of the 303 fully defined nucleotide sequences given in the  
 CC specification. The polypeptide is useful in the manufacture of a  
 CC medicament for treating a syndrome associated with a human disease. The  
 CC polypeptide, polynucleotide and antibody are useful in diagnosis.  
 CC treating or preventing NOVX-associated disorders, e.g. cancer, cachexia,  
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious  
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are  
 CC further used as hybridisation probes, in chromosome mapping, tissue  
 CC typing, preventive medicine, and pharmacogenomics. The present invention  
 CC represents a NOVX polypeptide of the invention.  
 CC  
 XX  
 SQ Sequence 2612 AA:  
 Query Match 77.4%; Score 13142.5; DB 8; Length 2612;  
 Best Local Similarity 91.6%; Pred. No. 0;  
 Matches 2457; Conservative 20; Mismatches 57; Indels 147; Gaps 12;

Db 421 HVARLEFQTDHSTGKRGFNITFTFRHNECPDGPVNVGKRFGBDSLQAGSSISFLCDEGF 480  
 QY 481 LGTQSGSTTTCVLKESGVVWNSAVLCEACGGHITSPSSTIISPGKPGFYKALSCAW 540  
 Db 481 LGTQSGSTTTCVLKESGVVWNSAVLCEACGGHITSPSSTIISPGKPGFYKALSCAW 540  
 QY 541 IEAOPGPXITPDRFTEVNVYTLLEKDRRTYSAPLIGVYHQVQPFISTSNLYL 600  
 Db 541 IEAOPGPXITPDRFTEVNVYTLLEKDRRTYSAPLIGVYHQVQPFISTSNLYL 600  
 QY 601 FSTDKSHSDIGQLRYETITLQSDHCLDPCIPVNGGHHNDPFGVALVTFSCSGYTLSD 660  
 Db 601 FSTDKSHSDIGQLRYETITLQSDHCLDPCIPVNGGHHNDPFGVALVTFSCSGYTLSD 660  
 QY 661 GEPLCECPNFWNSRALPSCGALCGFTQSSGHTLSPFDFPNNLNCWIIETSHGK 720  
 Db 661 GEPLCECPNFWNSRALPSCGALCGFTQSSGHTLSPFDFPNNLNCWIIETSHGK 720  
 QY 721 VFFTFHTFHLESQHDYLLITENGSPFOPLRLQSLRPLAPISAGLYGNFTAQRFIIDS 780  
 Db 721 VFFTFHTFHLESQHDYLLITENGSPFOPLRLQSLRPLAPISAGLYGNFTAQRFIIDS 780  
 QY 781 MSYEGFNITPSEYDLCEPEEVPAYSIRKGLQFVGDTLTFSCFPGYRLEGTAARITCLG 840  
 Db 781 MSYEGFNITPSEYDLCEPEEVPAYSIRKGLQFVGDTLTFSCFPGYRLEGTAARITCLG 840  
 QY 841 GRRRLWSSPLPRCAVEGNSVTGTGTLSPNPNVNNNHECTYSIQTOPGKIQIKAR 900  
 Db 841 GRRRLWSSPLPRCAVEGNSVTGTGTLSPNPNVNNNHECTYSIQTOPGKIQIKAR 900  
 QY 901 AFELSEBDYLVKVDGNNSARLLGVFSHSMGVTLNSTSSLMIDPITDAENTSKGFL 960  
 Db 901 AFELSEBDYLVKVDGNNSARLLGVFSHSMGVTLNSTSSLMIDPITDAENTSKGFL 960  
 QY 961 HFSSFELIKCEDPTKPFYKVDHGHFAGSSVFSQCDPYSLRGSEBLLCSGERRTWD 1020  
 Db 961 HFSSFELIKCEDPTKPFYKVDHGHFAGSSVFSQCDPYSLRGSEBLLCSGERRTWD 1020  
 QY 1021 RPLPTCYAECGGTVRGEVSGQVTSFGYPAHYENHNCITWTEBAGCTTGLHLVPTTEE 1080  
 Db 1021 RPLPTCYAECGGTVRGEVSGQVTSFGYPAHYENHNCITWTEBAGCTTGLHLVPTTEE 1080  
 QY 1081 VHDVLRIMDGPVBSGVLLKELSGPALPKDASHFNSVLDFTFTSKGFIQFSVST 1140  
 Db 1081 VHDVLRIMDGPVBSGVLLKELSGPALPKDASHFNSVLDFTFTSKGFIQFSVST 1140  
 QY 1141 ATSCNDPGIIPONSRSRGSWEAGDSTVFQCDPGYALQGSALISCVKIEKRFNFPQSPPTC 1200  
 Db 1141 ATSCNDPGIIPONSRSRGSWEAGDSTVFQCDPGYALQGSALISCVKIEKRFNFPQSPPTC 1200  
 QY 1201 IAPCGDLTSPSGVILSPNYPPEYPPGKECDKRTVVSDDYVIALVFNIFLBEQYFLHI 1260  
 Db 1201 IAPCGDLTSPSGVILSPNYPPEYPPGKECDKRTVVSDDYVIALVFNIFLBEQYFLHI 1260  
 QY 1261 YDGRDLSPLIGSFYSGQLPGRTESSNSLFLFRSPASVSNAGFVLDYENPRESCEFDP 1320  
 Db 1261 YDGRDLSPLIGSFYSGQLPGRTESSNSLFLFRSPASVSNAGFVLDYENPRESCEFDP 1320  
 QY 1321 GSINKGTRVSGDLKSSVTVYCHGGYEVGTLSCILPGDKPVNNNRPVCTAPCGG 1380  
 Db 1321 GSINKGTRVSGDLKSSVTVYCHGGYEVGTLSCILPGDKPVNNNRPVCTAPCGG 1380  
 QY 1381 QYVSGDGVILSPNYPONTYSGOICLVFTVYPKDYVVFQGFPAFHTALNDVEVHDGSHQ 1440  
 Db 1381 QYVSGDGVILSPNYPONTYSGOICLVFTVYPKDYVVFQGFPAFHTALNDVEVHDGSHQ 1440  
 QY 1441 SRLTSLISGHTGESLPLATSNQVLLKFSKGLAPARGFFVYQAVPRSATCCSVPEP 1500  
 Db 1441 SRLTSLISGHTGESLPLATSNQVLLKFSKGLAPARGFFVYQAVPRSATCCSVPEP 1500  
 QY 1501 RYGRKLSDPSFGVAIVRFECSNGYALQGSBEICLPPVGAALQMNVSAPTCVPPCGGNTL 1560  
 Db 1501 RYGRKLSDPSFGVAIVRFECSNGYALQGSBEICLPPVGAALQMNVSAPTCVPPCGGNTL 1560

QY 1561 ERRGIIISPGFPEPIYNSLNCVMKIVPEGAGIOIOVVSFVTEBQNDPMSLEVPDQADNTVT 1620  
 DB 1561 ERRGIIISPGFPEPIYNSLNCVMKIVPEGAGIOIOVVSFVTEBQNDPMSLEVPDQADNTVT 1620  
 QY 1621 MLGSPSGTIVALLNSTNSQLYLHPYSDISVSAAGFHLKYKTVGLSSCPBPAPVNSGVT 1680  
 DB 1621 MLGSPSGTIVALLNSTNSQLYLHPYSDISVSAAGFHLKYKTVGLSSCPBPAPVNSGVT 1680  
 QY 1681 GERLYVNDVVSFOCEPGYALOGHAHISCMPTVRNMYPPPLCTAOCGGVTEMEGIIIS 1740  
 DB 1681 GERLYVNDVVSFOCEPGYALOGHAHISCMPTVRNMYPPPLCTAOCGGVTEMEGIIIS 1740  
 QY 1741 PGFPGNTPSNMDCSMKIALPVFGAHIQPLNFSTEPNHDYIEIRNGEYETSRMMGRPSGS 1800  
 DB 1741 PGFPGNTPSNMDCSMKIALPVFGAHIQPLNFSTEPNHDYIEIRNGEYETSRMMGRPSGS 1800  
 QY 1801 ELPSLSLSTSHETTVYFHSHSQNRPPFKLEYQAYELQECPPDPBPANGIVRAGVNVQ 1860  
 DB 1801 ELPSLSLSTSHETTVYFHSHSQNRPPFKLEYQAYELQECPPDPBPANGIVRAGVNVQ 1860  
 QY 1861 SVTECLPGYOLTGHPVLTGCOHGNRMNDHPLPKCEVCGGNTSSNGTVSPSPSPYS 1920  
 DB 1861 SVTECLPGYOLTGHPVLTGCOHGNRMNDHPLPKCEVCGGNTSSNGTVSPSPSPYS 1920  
 QY 1921 SSQDCVWLITVPIGHGVRNLSLLQTEPSGDFITIMDGPQOTAPRLGVFTSRMAKTVOS 1980  
 DB 1921 SSQDCVWLITVPIGHGVRNLSLLQTEPSGDFITIMDGPQOTAPRLGVFTSRMAKTVOS 1980  
 QY 1981 SSNOVLKPHRDATAIGIPAIASAVPLTKCPPTILIPNAEVTNENEFIGIYVRCL 2040  
 DB 1981 SSNOVLKPHRDATAIGIPAIASAVPLTKCPPTILIPNAEVTNENEFIGIYVRCL 2040  
 QY 2041 PGFTLVGNEILLTKLGTYLOFEGEPPICEVHCPTNELITDSTGVISOSYPSGSPQOTC 2100  
 DB 2041 PGFTLVGNEILLTKLGTYLOFEGEPPICEVHCPTNELITDSTGVISOSYPSGSPQOTC 2100  
 QY 2101 SMLVREPDYNISLTYEYFLSEKQYDDEFIDPSPGQSLKALSGNYSAPLIVTSSNS 2160  
 DB 2101 SMLVREPDYNISLTYEYFLSEKQYDDEFIDPSPGQSLKALSGNYSAPLIVTSSNS 2160  
 QY 2161 VYLWMSDHAYNKRGFIRIRSAFYCSLPRALHGFILGQSTQPGSIHFCNAGRLVG 2220  
 DB 2161 VYLWMSDHAYNKRGFIRIRSAFYCSLPRALHGFILGQSTQPGSIHFCNAGRLVG 2220  
 QY 2221 HSMALCTTRHPQYHLSSEALPLCOALSCGLPEAPKXGMEYKTYVTGTAVYSCSEGYH 2280  
 DB 2221 HSMALCTTRHPQYHLSSEALPLCOALSCGLPEAPKXGMEYKTYVTGTAVYSCSEGYH 2280  
 QY 2281 QAGAEATAECLDTGMSNRNVPQCVPTCPDVSGISVHGMRLLIETQYQOQOLMI 2340  
 DB 2281 QAGAEATAECLDTGMSNRNVPQCVPTCPDVSGISVHGMRLLIETQYQOQOLMI 2340  
 QY 2341 CDPGYVYTGQVRVRCQANGKMSLGDSTPTCRITISCELPINPNHRIIGTLYGATAFIS 2400  
 DB 2341 CDPGYVYTGQVRVRCQANGKMSLGDSTPTCRITISCELPINPNHRIIGTLYGATAFIS 2400  
 QY 2401 CUSGYTLVSRVRECHANGMSGSEVRCLAGHGTPEPIVNHINGENYSYSGVYVQCN 2460  
 DB 2401 CUSGYTLVSRVRECHANGMSGSEVRCLAGHGTPEPIVNHINGENYSYSGVYVQCN 2460  
 QY 2461 AGFRLIGMSVRIQQODHHSKGTTPC-VPIITGCHGNPNVNGLTGQNFVNDVYKVCNP 2519  
 DB 2461 AGFRLIGMSVRIQQODHHSKGTTPC-VPIITGCHGNPNVNGLTGQNFVNDVYKVCNP 2519  
 QY 2519 GYMAEGARSOCLASQMSDMLPTCRI-----INCTDPGHQENSVRQV- 2562  
 DB 2519 GYMAEGARSOCLASQMSDMLPTCRI-----INCTDPGHQENSVRQV- 2562  
 QY 2562 -----HASGHRSPSGTIVSY-----RCN-----HGFYLLGTFTVLSQ 2595  
 DB 2562 -----HASGHRSPSGTIVSY-----RCN-----HGFYLLGTFTVLSQ 2595  
 QY 2595 DTKKIIVASTPAARTGSASAPLCLTGATTASTSWAPQCSAREMAHG--TVPASVSVS 2546  
 DB 2595 DTKKIIVASTPAARTGSASAPLCLTGATTASTSWAPQCSAREMAHG--TVPASVSVS 2546

QY 2596 GGGTWDRPRRPOCL-----LVSCGFGSPRPHSQMSGDSITYVG 2631  
 DB 2547 SGVTLWPSGLPASTLPDVWRQLYCGSSGA---VOLHROAYSGG 2584  
 RESULT 11  
 ID ADH72216  
 ADH72216 standard; protein, 3567 AA.  
 XX  
 AC ADH72216;  
 DT 25-MAR-2004 (first entry)  
 XX  
 DE Human protein of the invention NOV54b SEQ ID NO:1112.  
 XX  
 KW human; cytostatic; immunomodulator; neuroprotective; nootropic;  
 KW anorectic; antidiabetic; antimicrobial; antidiabetic; gene therapy;  
 KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;  
 KW obesity; diabetes; infectious disease; metabolic syndrome X;  
 KW dyslipidaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003102155-A2.  
 XX  
 PD 11-DEC-2003.  
 XX  
 PF 03-JUN-2003; 2003WO-US017430.  
 XX  
 PR 03-JUN-2002; 2002US-0385120P.  
 PR 04-JUN-2002; 2002US-0385784P.  
 PR 05-JUN-2002; 2002US-0386041P.  
 PR 05-JUN-2002; 2002US-0386047P.  
 PR 06-JUN-2002; 2002US-0386375P.  
 PR 06-JUN-2002; 2002US-0386435P.  
 PR 06-JUN-2002; 2002US-0386844P.  
 PR 06-JUN-2002; 2002US-0387016P.  
 PR 07-JUN-2002; 2002US-0386796P.  
 PR 07-JUN-2002; 2002US-0386816P.  
 PR 07-JUN-2002; 2002US-0386931P.  
 PR 07-JUN-2002; 2002US-0386942P.  
 PR 07-JUN-2002; 2002US-0386971P.  
 PR 07-JUN-2002; 2002US-0387262P.  
 PR 08-JUN-2002; 2002US-0296960P.  
 PR 10-JUN-2002; 2002US-0387400P.  
 PR 10-JUN-2002; 2002US-0387535P.  
 PR 11-JUN-2002; 2002US-0387610P.  
 PR 11-JUN-2002; 2002US-0387625P.  
 PR 11-JUN-2002; 2002US-0387634P.  
 PR 11-JUN-2002; 2002US-0387668P.  
 PR 11-JUN-2002; 2002US-0387668P.  
 PR 11-JUN-2002; 2002US-0387702P.  
 PR 11-JUN-2002; 2002US-0387836P.  
 PR 11-JUN-2002; 2002US-0387859P.  
 PR 12-JUN-2002; 2002US-0387933P.  
 PR 12-JUN-2002; 2002US-0387934P.  
 PR 12-JUN-2002; 2002US-0387960P.  
 PR 12-JUN-2002; 2002US-0388022P.  
 PR 12-JUN-2002; 2002US-0388096P.  
 PR 13-JUN-2002; 2002US-0389123P.  
 PR 14-JUN-2002; 2002US-0389118P.  
 PR 14-JUN-2002; 2002US-0389120P.  
 PR 14-JUN-2002; 2002US-0389144P.  
 PR 14-JUN-2002; 2002US-0389146P.  
 PR 17-JUN-2002; 2002US-0389729P.  
 PR 17-JUN-2002; 2002US-0389742P.  
 PR 18-JUN-2002; 2002US-0389884P.  
 PR 19-JUN-2002; 2002US-0390006P.  
 PR 19-JUN-2002; 2002US-0390209P.  
 PR 21-JUN-2002; 2002US-0390763P.  
 PR 17-JUN-2002; 2002US-0396706P.  
 PR 06-AUG-2002; 2002US-0401628P.  
 PR 09-AUG-2002; 2002US-0402156P.

PR 09-AUG-2002; 2002US-0402256P.  
PR 12-AUG-2002; 2002US-0402389P.  
PR 12-AUG-2002; 2002US-0402786P.  
PR 12-AUG-2002; 2002US-0402816P.  
PR 12-AUG-2002; 2002US-0402821P.  
PR 12-AUG-2002; 2002US-0402832P.  
PR 13-AUG-2002; 2002US-0403448P.  
PR 13-AUG-2002; 2002US-0403459P.  
PR 13-AUG-2002; 2002US-0403511P.  
PR 13-AUG-2002; 2002US-0403532P.  
PR 13-AUG-2002; 2002US-0403563P.  
PR 13-AUG-2002; 2002US-0406317P.  
PR 15-AUG-2002; 2002US-0406317P.  
PR 26-AUG-2002; 2002US-0406182P.  
PR 26-AUG-2002; 2002US-0406355P.  
PR 27-AUG-2002; 2002US-0406240P.  
PR 12-SEP-2002; 2002US-0412528P.  
PR 20-SEP-2002; 2002US-0412528P.  
PR 23-SEP-2002; 2002US-0412731P.  
PR 30-SEP-2002; 2002US-0414801P.  
PR 30-SEP-2002; 2002US-0414839P.  
PR 30-SEP-2002; 2002US-0414840P.  
PR 30-SEP-2002; 2002US-0414954P.  
PR 09-OCT-2002; 2002US-0417186P.  
PR 09-OCT-2002; 2002US-0417406P.  
PR 23-OCT-2002; 2002US-0420639P.  
PR 28-OCT-2002; 2002US-0421156P.  
PR 31-OCT-2002; 2002US-0422690P.  
PR 01-NOV-2002; 2002US-0423130P.  
PR 05-NOV-2002; 2002US-0423798P.  
PR 05-NOV-2002; 2002US-0423798P.  
PR 12-NOV-2002; 2002US-0425453P.

(CURA-) CURAGEN CORP.

PI Albrook JF, Alvarez E, Anderson DM, Boldog FE, Caeman SJ;  
PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;  
PI Ettenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;  
PI Gusev VY, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR;  
PI MacLachlan T, Malyanekar UM, Mezick AJ, Millet I, Mishra VS;  
PI Padigaru M, Paturajan M, Pena CE, Peyman JA, Raha D, Raetzelli L;  
PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shimkets RA;  
PI Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;  
PI Zhong H;

WPI; 2004-081935/08.  
DR N-PSDB; ADH72215.

PT New NOXV polypeptides and nucleic acid molecules useful for preventing or  
PT treating NOXV-associated disorders, e.g. cancer, diabetes, infection or  
PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.

XX Example 54; SEQ ID NO 112; 1880bp; English.

CC The invention relates to a novel isolated polypeptide (NOXV). A  
CC polypeptide of the invention has cytostatic, immunomodulator,  
CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and  
CC antiparasitic activity, and may have a use in gene therapy, and as a  
CC vaccine. The polypeptides are encoded by NOXV polynucleotides comprising  
CC any of the 303 fully defined nucleotide sequences given in the  
CC specification. The polypeptide is useful in the manufacture of a  
CC medicament for treating a syndrome associated with a human disease. The  
CC polypeptide, polynucleotide and antibody are useful in diagnosing,  
CC treating or preventing NOXV-associated disorders, e.g. cancer, cachexia,  
CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious  
CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are  
CC further used as a hybridisation probe, in chromosome mapping, tissue  
CC typing, preventive medicine, and pharmacogenomics. The present sequence  
CC represents a NOXV polypeptide of the invention.

XX Sequence 3567 AA;

Query Match 66.2%; Score 11247.5; DB 8; Length 3567;

Best Local Similarity 63.6%; Pred. No. 0;  
Matches 1957; Conservative 458; Mismatches 600; Indels 61; Gaps 11;  
QY 1 MAGAPPALI-----LPCSLISD-----CCASQNRHSVGVPSSELYKKQIELKSRGV 47  
Db 264 ISGTEAPSIWLTGMNLPSPVSSKMLRLHFTSDSNRRK-GPNAQFOVKKAIELKSRGV 322  
QY 48 KLPSPKDNQKTSVLTQVGVSGAHMCCPDGIERGKRGISDP-RLGSSVQFCTNEDYVL 106  
Db 323 KMLPSKDSKHSVLSQGVALVSHMCLDPIENGRRASDSRVGAVNQFSCEDYVL 382  
QY 107 QGSKRITCMKVSMDFAWSDHRRPVCRAHMDAHLRGPSSGIIITSPNPIQYNNAHCVI 166  
Db 383 QGSKSITCQVETETLAAMSHRPICARITGSLNLRGSGVITSPNPIQYNNAHCVI 442  
QY 167 TALNPKVITLAFEEFDLERGYDTLVGDGGQDQKTVLYMSQACSDSPHTPGRIPE 226  
Db 443 TTDPDKVITLAFEEFDLERGYDTLVGDGAKGDKTRSVLYV-----LTGSSVPD 492  
QY 227 ---SMGDIWROKMTVLEICRDISSSDARSGVRSKPSKTSNAVELVAPG-----TEEOG 278  
Db 493 LIVSMNSQWMLH-----LQSDS-----ISPGKRAYQIEK 526  
QY 279 SCGDPIGIPAYGRREGSRFHFGDTLKFECPAPAFELVGOIKATCOKNQWSAKKPGCVPSCF 338  
Db 527 GCGDPGIPAYGKRTGSSFLHGDITLFECPAPAFELVGERVITCOQNNQWSGKPSCVPSCF 586  
QY 339 FNFTSPGVVLSBNPYEDYGNHLHCVLLIAREBSRIHLAFNDI DVEPQDFLVIXDGT 398  
Db 587 FNFTASGIIISBNPYEYGNMNCVLLIAREBSRIHLAFNDI DVEPQDFLVIXDGT 646  
QY 399 AEAIVLGTFSGNOLPSSITSSGHVAPLEPOTDSTGKRGNIITFTFRHNECDPGVNV 458  
Db 647 SDITVLTGFSGNVNPQLASSGHI VLEPFDSDSTGKRGNIITFTFRHNECDPGVNV 706  
QY 459 GKRFDSLIQSSISIFLACDGGFLGTGSEITITVLKESGVNMSAVLRCAPCGHILSP 518  
Db 707 GRFGDRFLGSSVSFHCDDGFKTQSSBITCLIDGNVMSSTVRCAPCGHILSP 766  
QY 519 SGTILSPGWFYKDALSCAMVIEAOPGYIKTTPDKTEVNYDTLEVDGRTYSAPLI 578  
Db 767 SGVILPGMGVYKDSIHCMIIEAKGHSIKMTDFQFEVNVDTLEVDGRTYSAPLI 826  
QY 579 GYHGTQVPOPLISTSVLYLFTSTKSHSDIFOLRYETITLOSDICLDPGIPVNGQRH 638  
Db 827 GEYHGTQVPOPLISTSVLYLFTSTKSHSDIFOLRYETITLOSDICLDPGIPVNGQRH 886  
QY 639 GNDPVGALVTFSCDSGYTLDGSEPLECEPNFQMSRALPSCALCGPIQSSGIIITSPG 698  
Db 887 GGFPGISITVTFSCDPSYITLSDDEPLVCERNQNMHALPSCDALCGVIGKGTIVLSPG 946  
QY 699 FPDYFNNLNCWTIETSHGKGVFTPTHTLSSGHDYLLITENGSTFQPLROLTGSRLP 758  
Db 947 FPDYFNNLNCWTIETSHGKGVFTPTHTLSSGHDYLLITENGSTFQPLROLTGSRLP 1006  
QY 759 APIASGIXGFTQVRIISFMSYEGFNITTSBYDLEPEEBEPVAPVYSTRKGLQFVGD 818  
Db 1007 HTTKAGIFGFTQVRIISFMSYEGFNITTSBYDLEPEEBEPVAPVYSTRKGLQFVGD 1066  
QY 819 TLFPSCPGVRLBGTARTITLGGRRRLMSPLPRCVACGNSVTGCTILSPFNPVNV 878  
Db 1067 SLTFSCVGLRLBGTARTITLGGRRRLMSPLPRCVACGNSVTGCTILSPFNPVNV 1126  
QY 879 NNHECIYSIOTOPKGIQLKARAFELSEGVLVKYDGNNSARLLGVFSHEMVGTVLNS 938  
Db 1127 NTHCEIKYIRTEAGKGIHLTRSPQLPEGDTLKVYDGDSSRPLGFTFXNELLGLTINS 1186  
QY 939 TSSSLMIDFTTDAENSKGFLHPSFELKCDPGRPKGYKVDHGHPAGSSVSFSCD 998  
Db 1187 TSNHMLLEFNTNSGDDTQGFQLTYSFDLVKCDPGRPKGYKVDHGHPAGSSVSFSCD 1246  
QY 999 PGVSLRSGSEILCSGERRTMDRPLPTCVACGCTVAGVSGOVLSPGYAPAEHNLNCI 1058

Db 1247 PGYAMHSGNTLTCTLSGDRRWMDKPLPSCIAECGQIHAATSGRIILSPGYAPYDNNLHCT 1306  
Qy 1059 WTEEAAGCTIGLHETVPTTEBHVDLRLTWDPVBSGVLLKELSGPALPKDLASTFNSV 1118  
Db 1307 WTEEAAPGKTIISHFTVPTTEMAHDLTKWMDGPVSDILLKEMSSGALPEDHSTFNSLT 1366  
Qy 1119 LQFSTDFPSSKOGFAIQFSSVSTATSCNDPGI PONSRSRSGDSWEADSDVTFQCDPEYALOG 1118  
Db 1367 LQFSDDFPSSKOGFISIQFSTSIATCNDGMPQNGTRGDSSEADDTVTFQCDPEYALOG 1426  
Qy 1179 SAEISGVKLENFPFQOPSPPTCIAFCGDLTPSGVILSPNPEPYPKGECMDKVTVP 1238  
Db 1427 QAKITCVOLMNRFPQOPBPPTCIAACGGLTGPAAVILSPNPGPYPKGECMDKVTVP 1486  
Qy 1239 DVYALVFNILFLEPGYDFLHYDGRDLSPLISFGYSOLPGRLESSNLSFLAFPSDA 1238  
Db 1487 DVYALVLFKSPMESPSTDLHYEGEDSNPLISGYSQAPBERLESSGNLSFLAFPSDA 1546  
Qy 1299 SVSNAGFYVDYENPRESCEPDGSIKNGTRVSGDLKSGSVYVYCHGGVEVEGTSTLSGI 1358  
Db 1547 SVGLSGFALEFKEKEREACFPDGNIMNGTRVGTDPKLGSTITTCDSGKILDPSSITCV 1606  
Qy 1359 LGPDKPVMNNRPVCTAPCGGQVYSGDVLSPPYQNTYSGQICLYFVTVPKDYVFG 1418  
Db 1607 IGADKPSMDVYLPSCNAPCGQYTGSEGVLSPPYNNYTAQOICLYSITVPEKPFVVG 1666  
Qy 1419 QPAFHTALNDVNEVHDHSHSRLLSSLSGSHTGESLPLATSNQVLIKFSAKGLAPARG 1478  
Db 1667 QPAFHTALNDVNEVHDHSHSRLLSSLSGSHTGESLPLATSNQVLIKFSAKGLAPARG 1726  
Qy 1479 FHFYVQAPRTSAQOCSSVPEPRYKRLGSDPSGAIIVREPCNSGVALGSSPETECLPVP 1538  
Db 1727 FHFYVQAPRTSDQOCSSVPEPRYKRLGSDPSGAIIVREPCNSGVALGSSPETECLPVP 1786  
Qy 1539 GALAQMNVASPTCVVPCGGLTERGTILSPGFPEPLNSCWNKIVVEGAGIQOVV 1598  
Db 1787 NALAQMNDITPSCVVPCGGLTERGTILSPGFPEPLNSCWNKIVVEGAGIQOVV 1846  
Qy 1599 SFVTEQWMD-SLEFDPQADNTVMLGSPGTTVPALINSTNOLYHFSYDISVSAAGHP 1657  
Db 1847 QFCHGABLGHPFELHHDGDVYAPRLGFSGTVPALINSTNOLYHFSYDISVSAAGHP 1906  
Qy 1658 LEYKTVGLSSCEPEPAVNSGVKTGERLYLVNDVVSFOCEPGYALOGMAHISCMPTGRAN 1717  
Db 1907 LEYKTVGLSSCEPEPAVNSGVKTGERLYLVNDVVSFOCEPGYALOGMAHISCMPTGRAN 1966  
Qy 1718 YPPPLCIAQCGGVTEENEGVILSPGFPGNPSNMDCSMKIALPVGCAHIQELNFTSTEN 1777  
Db 1967 YPPPLCIAQCGGVTEENEGVILSPGFPGNPSNMDCSMKIALPVGCAHIQELNFTSTEN 2026  
Qy 1778 HDYIEIRNGPYETSRMMGRFSGSELBSLSTSHETTVYFHSHSQNRPGFLEYQAYEL 1837  
Db 2027 HDYIEIRNGPYETSRMMGRFSGSELBSLSTSHETTVYFHSHSQNRPGFLEYQAYEL 2086  
Qy 1838 QECPEPFPANIVRGAGVNGSVTPECLPGYOLGHPVLTCOHGNTNNWHPYKPCV 1897  
Db 2087 QECPEPFPANIVRGAGVNGSVTPECLPGYOLGHPVLTCOHGNTNNWHPYKPCV 2146  
Qy 1898 PCGNTTSSNGTVSSGFPSPVSSSODCWLITVPIGHVRLNLSLQTEPSGDEITIMD 1957  
Db 2147 PCGNTTSSNGTVSSGFPSPVSSSODCWLITVPIGHVRLNLSLQTEPSGDEITIMD 2206  
Qy 1958 GPOQTAPRLGVFTSRMAKKTVOSSNQVLLKFRBDAATGIFALFSAVPLTKCPTTIL 2017  
Db 2207 GPOQTAPRLGVFTSRMAKKTVOSSNQVLLKFRBDAATGIFALFSAVPLTKCPTTIL 2266  
Qy 2018 PNAEVTEBENIGDIVRKLPGFTLVGNEILTKLGTYLQFEGPPICSVHCTNEL 2077  
Db 2267 PNAEVTEBENIGDIVRKLPGFTLVGNEILTKLGTYLQFEGPPICSVHCTNEL 2326  
Qy 2078 LFDSTGVILSOGSPGSYFQFOTCSMLVREPPYINSLVVEYFLSKQYDFEPIFDGSPGQ 2137  
Db 2327 LFDSTGVILSOGSPGSYFQFOTCSMLVREPPYINSLVVEYFLSKQYDFEPIFDGSPGQ 2386

Qy 2138 SPLKALISGVASAPLIVTSSSNSVYLRMSSDHAVNKGFIRYASAPYCSLPAPLHFTIL 2197  
Db 2387 SPLVVLSSGHTGNSQNSPTSRNSQVLYRMSTDAHSTKKGKIRYAPAYCSLTHPLKNGGIL 2446  
Qy 2198 GQTSSTOPGSGIHFECNAGYRVLVGHSMALICTRHPOGYHLSSEALPLCOALSCGLPEAPKNG 2257  
Db 2447 NRTAGAVSGKWHFCEKPRGYRMVGHSAVATCRBNPLGMQVDSILPLCOAVSCGIPSPNG 2506  
Qy 2258 MVEKEVYTGIVAVVYSCSEGYHLOGAETACCLDTGLMSNNVNPQCYVPTCPDVSSIS 2317  
Db 2507 SFTGNEFTLDSKVYVECHGEFPLSSQQLTAVCOEDGLMSNKGKPMCKPVACPSIEQL 2566  
Qy 2318 VEHGRNRLIFETQYQFQALMLICDPGYVYTGQVYRICOANGKMSLGDSTPCRIISGE 2377  
Db 2567 SEHVIRLVSGSLNEGAQVLLSCGPYLLBEMRLLRQANCTWNIIGBERBCKRYSIGS 2626  
Qy 2378 LPIPMGHRIGTLVYAGATATFSGNSGTYLVGSRVRECMANGLSGSEVRCLAGHCTPE 2437  
Db 2627 LSPPPGNKIGTLVYAGATATFTCNTGYLVGSHVRECLANGLSGSEVRCLAGHCGSPD 2686  
Qy 2438 PIVNGHINENYSGSVVYOCNAGFRILIGMSVRIQOQDHHMSGKTPPCVPTTCHGPNP 2497  
Db 2687 PIVNGHISGDSFSYRDTVYOCNPGFRVLGTSVRIQOQDHHMSGKTPPCVPTTCHGPNP 2746  
Qy 2498 VNGLTQGNQNLNDVYKVCNPGYVABEGAAASQCLASQWSMDLPTCRINCTDPGHQEN 2557  
Db 2747 AHGFTNGSEFNLNDVNFCTCNTGYLVQGVSAQCRSNQWSMDLPTCRINCTDPGHQEN 2806  
Qy 2558 SVRQVHAAGPRHFSFGTIVSYRCHNGFYLCTPVLSQCGDCTMRPRPQCLLVSCGHPGS 2617  
Db 2807 AIRHQGNQPRHFSFGTIVSYRCHNGFYLCTPVLSQCGDCTMRPRPQCLLVSCGHPGS 2866  
Qy 2618 PPHSQMSDSTVGAIVYRSCIGRTLVGNSTRMCGLDGHWTGSLPHCSGTSVGVCDPG 2677  
Db 2867 PPHSQMSDSTVGAIVYRSCIGRTLVGNSTRMCGLDGHWTGSLPHCSGTSVGVCDPG 2926  
Qy 2678 IPAHGIRLGDSFPDGTWRSECEAGHYLRGSEERTCOANGMSGSOPECVYISCGNPGTP 2737  
Db 2927 IPAHGIRLGDSFPDGTWRSECEAGHYLRGSEERTCOANGMSGSOPECVYISCGNPGTP 2986  
Qy 2738 SNARVPSDGLVFSSSIVYECREGEVYATGLSHRCSVNGTWTGSDPECLVYINCDDPGIPA 2797  
Db 2987 SNARVPSDGLVFSSSIVYECREGEVYATGLSHRCSVNGTWTGSDPECLVYINCDDPGIPA 3046  
Qy 2798 NGALGDNFRKNTVYQVCPGYMMESHRYVSLCTKQRTWNGTKPVCKALMKCPPLIP 2857  
Db 3047 NGALGDNFRKNTVYQVCPGYMMESHRYVSLCTKQRTWNGTKPVCKALMKCPPLIP 3106  
Qy 2858 NGKTVGSDPFMWSGVYACLEGYQSLPAYFTCEGNGSWTGELOPCEPVPCGPGVPSRG 2917  
Db 3107 NGKTVGSDPFMWSGVYACLEGYQSLPAYFTCEGNGSWTGELOPCEPVPCGPGVPSRG 3166  
Qy 2918 RREBRGYSYSSVFSCHPPLVIVGSPRRFCQSDGTWSGTQPSCI-DPTLTTCADPGVPQ 2976  
Db 3167 RREBRGYSYSSVFSCHPPLVIVGSPRRFCQSDGTWSGTQPSCI-DPTLTTCADPGVPQ 3226  
Qy 2977 FGIQNSOGQVQSYLVFRCQKGYLLQGSTTRCLDNLWMSGTRPPCVPHRCQPTPTPH 3036  
Db 3227 FGIQNSOGQVQSYLVFRCQKGYLLQGSTTRCLDNLWMSGTRPPCVPHRCQPTPTPH 3286  
Qy 3037 ANYGALDLPBMGYTLI 3052  
Db 3287 ADVRAIDLPFGYTLV 3302

RESULT 12  
AAE20789 standard; protein; 3100 AA.  
AAE20789;  
AC  
XX  
XX  
01-JUL-2002 (first entry)



QY 1382 YVSGDVVLSPNYSPONTYSQOICLYPVTPEKDYVFGOPAFPHHALNDVVEVDHGSQHS 1441  
 DB 1165 YTGSEGVVLSPNYHNHTAGQICLXSTVPEKXVFGOPAFPHHALNDVVEVDHGSQHS 1224  
 QY 1442 RLSSLSGSHGTESLPLATSNQVLIKFSAKGLABARGFHYQAVPTSAOCSSVEPR 1501  
 DB 1225 RLSSLSGSHGTESLPLATSNQVLIKFSAKGLABARGFHYQAVPTSAOCSSVEPR 1284  
 QY 1502 YGKRLSGDESGALVREPCNSGYALQSPLECLPVFGALAQWVSAFTCVVPCGKALTE 1561  
 DB 1285 YGKRLSGDESGALVREPCNSGYALQSPLECLPVFGALAQWVSAFTCVVPCGKALTE 1244  
 QY 1562 RRGITLSPGPEPEPLANSNCWKIVPEAGAGIQOVASFTEQWMDSELDHGDQVTAAPR 1621  
 DB 1345 RRGITLSPGPEPEPLANSNCWKIVPEAGAGIQOVASFTEQWMDSELDHGDQVTAAPR 1404  
 QY 1622 LGSSGCTTVPALNLSNOLYHFDISYSAAGFHELYKTVGLSSCPREPAVPSNGKVG 1681  
 DB 1405 LGSSGCTTVPALNLSNOLYHFDISYSAAGFHELYKTVGLSSCPREPAVPSNGKVG 1464  
 QY 1682 ERYLVNDVSPQCPGVALQGHANISCMPTVRMNPPLCTAACCGTYEEMEGYILSP 1741  
 DB 1465 ERYLVNDVSPQCPGVALQGHANISCMPTVRMNPPLCTAACCGTYEEMEGYILSP 1524  
 QY 1742 GFGNVPNSMDCSKIALPVFGAHIOPLNFSTEPNHDYIEIRNGPEYETSRMMGRFSGE 1801  
 DB 1525 GFGNVPNSMDCSKIALPVFGAHIOPLNFSTEPNHDYIEIRNGPEYETSRMMGRFSGE 1584  
 QY 1802 LBSLSLSTSHETTVYFRSHDSQNPGRKLEYQAYEIOECDPPEPPANGIVRAGYVNGS 1861  
 DB 1585 LBSLSLSTSHETTVYFRSHDSQNPGRKLEYQAYEIOECDPPEPPANGIVRAGYVNGS 1644  
 QY 1862 VTECELPGYOJTGHPVLTQCHGTRNBMHPLPKCEVPCGNTISSNGTYVSPFEPYPS 1921  
 DB 1645 VTECELPGYOJTGHPVLTQCHGTRNBMHPLPKCEVPCGNTISSNGTYVSPFEPYPS 1704  
 QY 1922 SODCWMLITVPIGHVRLNLSLQTEPSGDFITIMDPOQTAERLGVFTSMMAKTVQSS 1981  
 DB 1705 SODCWMLITVPIGHVRLNLSLQTEPSGDFITIMDPOQTAERLGVFTSMMAKTVQSS 1764  
 QY 1982 SNOVLKFRHDATGCIFAIAESAYPLTKCPPTIIPMAEVTEENEFNIGDIYRCLP 2041  
 DB 1765 SNOVLKFRHDATGCIFAIAESAYPLTKCPPTIIPMAEVTEENEFNIGDIYRCLP 1824  
 QY 2042 GFTLVGNEILTCAGTYLOEPGPPICEVHCFPTNEILTSTGTYLSSQVSGSPQFQCS 2101  
 DB 1825 GFTLVGNEILTCAGTYLOEPGPPICEVHCFPTNEILTSTGTYLSSQVSGSPQFQCS 1884  
 QY 2102 WLVRVEDYINISLVEYFLSEKQYDEFEIJDGSGSPILKALSGNYSAPLIVTSSNSV 2161  
 DB 1885 WLVRVEDYINISLVEYFLSEKQYDEFEIJDGSGSPILKALSGNYSAPLIVTSSNSV 1944  
 QY 2162 YLRMSDHDVNRKGFIRYASAPYCSIPRAPHFLIGOSTOPGSHHGCVNAGRLVGH 2221  
 DB 1945 YLRMSDHDVNRKGFIRYASAPYCSIPRAPHFLIGOSTOPGSHHGCVNAGRLVGH 2004  
 QY 2222 SNAICTRHPQGYHLMSEAIPLCOALSCGPEAPKNGVNGKTYVTGTVKAVSCSEGYH 2281  
 DB 2005 SNAICTRHPQGYHLMSEAIPLCOALSCGPEAPKNGVNGKTYVTGTVKAVSCSEGYH 2064  
 QY 2282 AGAEATACLDJTGMLNSNRNVPPOCVPTCPDVSSISVEHGRWLIIFETQYQFOALMLIC 2341  
 DB 2065 AGAEATACLDJTGMLNSNRNVPPOCVPTCPDVSSISVEHGRWLIIFETQYQFOALMLIC 2124  
 QY 2342 DPGYVYTGQVIRCOANGKMSLGDSTPTCRITISCGELPPIPRGHRTIGTISVGAARISG 2401  
 DB 2125 DPGYVYTGQVIRCOANGKMSLGDSTPTCRITISCGELPPIPRGHRTIGTISVGAARISG 2184  
 QY 2402 NSGYTLVGRVRECMANGMLSGSEVYCLAGHGTPEPIYNGINGENYSYRGSVYQCN 2461  
 DB 2185 NSGYTLVGRVRECMANGMLSGSEVYCLAGHGTPEPIYNGINGENYSYRGSVYQCN 2244

QY 2462 GPELGMVRIQCOODHMSGKTPFCVPTTCGHPNPVNLTOGNQPNLDVVKFVCPNPGY 2521  
 DB 2245 GPELGMVRIQCOODHMSGKTPFCVPTTCGHPNPVNLTOGNQPNLDVVKFVCPNPGY 2304  
 QY 2522 MEGARSQCLASQGMWMLPTCRILINCTDPGHQENSVYQVASGPHRSFGTTSYRCN 2581  
 DB 2305 MEGARSQCLASQGMWMLPTCRILINCTDPGHQENSVYQVASGPHRSFGTTSYRCN 2364  
 QY 2582 HGFYLLGTIVLSQCGDGTDRPPOCLLVSCGHPSPPHSQNSGDSYTVGAVVYSCLK 2641  
 DB 2365 HGFYLLGTIVLSQCGDGTDRPPOCLLVSCGHPSPPHSQNSGDSYTVGAVVYSCLK 2424  
 QY 2642 RTLVGNTSMGLDGHVWGLPHCGSTVSGVGDGCIAPHGRLDSDPDPGVMRPSCA 2701  
 DB 2425 RTLVGNTSMGLDGHVWGLPHCGSTVSGVGDGCIAPHGRLDSDPDPGVMRPSCA 2484  
 QY 2702 GHVLRSSERTCOANGSMGSGQPEGVISQGNPTSPNARVVPDGLVFPSSIVYECREG 2761  
 DB 2485 GHVLRSSERTCOANGSMGSGQPEGVISQGNPTSPNARVVPDGLVFPSSIVYECREG 2544  
 QY 2762 YVATGLSRHSGVNGTWSGDECLVINGDGPGLPANGRLANDFRYKTVTYQCPGYM 2821  
 DB 2545 YVATGLSRHSGVNGTWSGDECLVINGDGPGLPANGRLANDFRYKTVTYQCPGYM 2604  
 QY 2822 MESHVSVLSTCKDRTMNGTKRVCKALCKPPELLPNKVVYSDPMGSSVYACLEGY 2881  
 DB 2605 MESHVSVLSTCKDRTMNGTKRVCKALCKPPELLPNKVVYSDPMGSSVYACLEGY 2664  
 QY 2882 LSLPAVFTCEGNGSTGELPOCFPVFCGDPGVPSRGRREDGFSYRSSVSFSCHPPLVY 2941  
 DB 2665 LSLPAVFTCEGNGSTGELPOCFPVFCGDPGVPSRGRREDGFSYRSSVSFSCHPPLVY 2724  
 QY 2942 GSPRRFQSDGWSGTOSTCIDLPTLTTCADRGVPOFGIQQNSQGYQVGSYTLFRCOKGYL 3001  
 DB 2725 GSPRRFQSDGWSGTOSTCIDLPTLTTCADRGVPOFGIQQNSQGYQVGSYTLFRCOKGYL 2784  
 QY 3002 LQGSTTRTCLPMLTWSGTPPCVPHRCQPEPTPTANGALDLPBMGTLLI 3052  
 DB 2785 LQGSTTRTCLPMLTWSGTPPCVPHRCQPEPTPTANGALDLPBMGTLLI 2835

RESULT 13  
 ID AAE20901 standard; protein; 3100 AA.  
 AC AAE20901;  
 DT 01-JUN-2002 (first entry)  
 DE Human C3b/C4b complement receptor like protein #2, alternative version.  
 KW Human; C3b/C4b complement receptor-like molecule; immune system disorder;  
 KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;  
 KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;  
 KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;  
 KW transplant rejection; autoimmune disease; ischemic condition; necrotic;  
 KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;  
 KW infertility; vasodilator; obesity; cardiac.  
 OS Homo sapiens.  
 XX  
 XX  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 418  
 FT /label= unknown  
 FT  
 FT Misc-difference 492  
 FT /label= unknown  
 FT  
 FT Misc-difference 617  
 FT /label= unknown  
 FT  
 FT Misc-difference 726  
 FT /label= unknown  
 FT  
 FT /note= "Encoded by CTR"  
 XX  
 PN W0200210199-A2.



QY 1742 GEPGVPSNMDCSKIALPVGFGAHIQIPLNFSTERNNDYIIRNGPYETSRMGRFEGSE 1801  
 DB 1525 GPPSGYVNNLDCITWRKISLPIDYGAHIQIPLNFSTERNNDYIIRNGPYETSRMGRFEGSE 1584  
 QY 1802 LPSLLSTSHETTYVFNHSDHSONRPGFYLEYQAYELOECPDPPEFPANGIVRAGAYNGOS 1861  
 DB 1585 LPAALSTTHETHTLHFFPSDHSQONQGFPLAQAELOECPDPPEFPANGIVRAGAYNGOS 1644  
 QY 1662 VTEFCLRGYQLTGHVPLTQCHGTRNNDHPIPKCEVPDGGNITSSNGTVYSPGFPDYS 1921  
 DB 1445 VSEFCYRGYLLIGHVPLTQCHGTRNNDHPIPKCEVPDGGNITSSNGTVYSPGFPDYS 1704  
 QY 1922 SODCVMLITVPIGHVPLTQCHGTRNNDHPIPKCEVPDGGNITSSNGTVYSPGFPDYS 1981  
 DB 1705 LKDCIMLITVPIGHVPLTQCHGTRNNDHPIPKCEVPDGGNITSSNGTVYSPGFPDYS 1764  
 QY 1982 SNOVLLKFRHDAATGIFAIAPSAVPLTKCPPTILPNAVEVNEEFNIGDIYRCLP 2041  
 DB 1765 TNQVLLKFRHDAATGIFAIAPSAVPLTKCPPTILPNAVEVNEEFNIGDIYRCLP 1824  
 QY 2042 GFTLVGNEILLTKCTGTLOFEGPPICEVHCPTNELLTDTGTYLLSGTYSGSPQOTCS 2101  
 DB 1825 GFTLVGNEILLTKCTGTLOFEGPPICEVHCPTNELLTDTGTYLLSGTYSGSPQOTCS 1884  
 QY 2102 MLYVREBDYNIISLVEYFLESEKQDEFEIFDGSGGSPLLKALSGVNSAPLIYSSNSV 2161  
 DB 1885 MSIKVENNYNITTFVDTFQSEKOPDALEVFDGSGGSPLLKALSGVNSAPLIYSSNSV 1944  
 QY 2162 YLRWSSDHANRKGFKIRYAPFCSLPRAPLHGFILQSTQSTPGSIFHECCNAGYRLVGH 2221  
 DB 1945 YLRWSSDHANRKGFKIRYAPFCSLPRAPLHGFILQSTQSTPGSIFHECCNAGYRLVGH 2004  
 QY 2222 SMAICTHPPGYHMLMSEAIPLCOALSCGLPEARKNVPEKKEYTGKAVYSGSEBHQ 2281  
 DB 2005 SNAICTHPPGYHMLMSEAIPLCOALSCGLPEARKNVPEKKEYTGKAVYSGSEBHQ 2064  
 QY 2282 AGABATACLDLGLMSNRNVPPOCVPTCPDVSISVEHGRMLIFETQYFOAQLMLIC 2341  
 DB 2065 SSOQATFVCOEDGLMSNRNVPPOCVPTCPDVSISVEHGRMLIFETQYFOAQLMLIC 2124  
 QY 2342 DPGYVYTGQVIRCOANGKSLDSTPTCTIISCGELPIPPNGHRIIGTISVGAIAIPSC 2401  
 DB 2125 SPQYVYTGQVIRCOANGKSLDSTPTCTIISCGELPIPPNGHRIIGTISVGAIAIPSC 2184  
 QY 2402 NSGYTVGSRVRECMANGLSGSEVRCLAGHCTPPIYNGHNGENYVRSGVVQCN 2461  
 DB 2185 NTGYTVGSRVRECMANGLSGSEVRCLAGHCTPPIYNGHNGENYVRSGVVQCN 2244  
 QY 2462 GFTLVGSRVRECMANGLSGSEVRCLAGHCTPPIYNGHNGENYVRSGVVQCN 2521  
 DB 2245 GFTLVGSRVRECMANGLSGSEVRCLAGHCTPPIYNGHNGENYVRSGVVQCN 2304  
 QY 2522 MARGASOCIASGOMSDMLPTCRILNCDPQHGENSVQVHASGHRSEFTGYVSYRN 2581  
 DB 2305 LUGVSRASOCIASGOMSDMLPTCRILNCDPQHGENSVQVHASGHRSEFTGYVSYRN 2364  
 QY 2582 HGFYLLGTPLVSCQGDGTWDRPRPQCLIVSCGHPGSPHSONSGSYTVGAVVRSCTGK 2641  
 DB 2365 KGFYLLGTPLVSCQGDGTWDRPRPQCLIVSCGHPGSPHSONSGSYTVGAVVRSCTGK 2424  
 QY 2642 RTLVGSRVRECMANGLSGSEVRCLAGHCTPPIYNGHNGENYVRSGVVQCN 2701  
 DB 2425 ESTLIGDTRVCOEDGLMSNRNVPPOCVPTCPDVSISVEHGRMLIFETQYFOAQLMLIC 2484  
 QY 2702 GHVLRSSERTCOANGSGOPEGVISCANPGTSPSNARVVFSSVSSIVVECEG 2761  
 DB 2485 GHVLRSSERTCOANGSGOPEGVISCANPGTSPSNARVVFSSVSSIVVECEG 2544  
 QY 2762 YVATGLSHRCSVNGTWSDEPCLVINGCDPQIPANGLRNDPRYNTVYQCVPGYM 2821  
 DB 2545 YVATGLSHRCSVNGTWSDEPCLVINGCDPQIPANGLRNDPRYNTVYQCVPGYM 2604

QY 2822 MESHRSVLSCTYDRTWNGTKPVCAALMCKPPPLIPNGRVGSDFMWSSSVTYACLEGY 2881  
 DB 2605 MEAVTSATIRCTYDGRMNSKEVCRAVLCOPPPVONGTVBESDFRMWSSSVTYACLEGY 2664  
 QY 2882 LSPAVFTCEGNSWNGELPOCFPPVCCGPGVPSNGRREDRFSTRSSVSFSCHPPLVY 2941  
 DB 2665 LSHSALISCEGNSWNGELPOCFPPVCCGPGVPSNGRREDRFSTRSSVSFSCHPPLVY 2724  
 QY 2942 GSPRRPQSDGTWSTGTPSCIDPTLTTCADPGVPOFGIOMNSQGVQVSTVLPFCOKGYL 3001  
 DB 2725 GSPRRPQSDGTWSTGTPSCIDPTLTTCADPGVPOFGIOMNSQGVQVSTVLPFCOKGYL 2784  
 QY 3002 LOGSTRTCLPULTWSGTPPDVCPVHRCROPEFTTANALDLPNGTYLLI 3052  
 DB 2785 LOGSTRTCLPULTWSGTPPDVCPVHRCROPEFTTANALDLPNGTYLLI 2835

## RESULT 14

AAE20788 standard; protein; 3095 AA.

AAE20788;

01-JUL-2002 (first entry)

Rat C3b/C4b complement receptor like protein.

KW Rat; C3b/C4b complement receptor-like molecule; immune system disorder;  
 KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;  
 KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;  
 KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;  
 KW transplant rejection; autoimmune disease; ischemic condition; nocturnal;  
 KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;  
 KW infertility; vasodilator; obesity; cardiant.

Rattus rattus.

Key Location/Qualifiers

FT Misc-difference 196

FT /label= "Unknown"

FT /note= "Xaa can be any amino acid"

FT /label= "Unknown"

FT /note= "Xaa can be any amino acid"

PN WO200210199-A2.

PD 07-FEB-2002.

PF 24-JUL-2001; 2001WO-US023232.

PR 02-AUG-2000; 2000US-0222504P.

PR 28-NOV-2000; 2000US-00728787.

XX (AMGE-) AMGEN INC.

XX PI Welcher AA, Elliott GS;

XX WPI, 2002-303934/34.

XX N-PSDB; AAD33319.

XX Nucleic acid encoding a novel C3b/C4b complement receptor-like nucleic

XX acid molecule, useful for treating, preventing and diagnosing rheumatoid

XX arthritis, psoriatic arthritis, inflammatory arthritis, and multiple

XX sclerosis.

XX Claim 13; Fig 3; 251pp; English.

XX The invention relates to a nucleic acid encoding a novel C3b/C4b

XX complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like

XX polypeptide and nucleic acid molecules may be used to treat, prevent,

XX ameliorate, diagnose and/or detect diseases such as immune system

XX disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory

CC arthritis, osteoarthritis, inflammatory joint disease, autoimmune  
 CC disease, multiple sclerosis, lupus, inflammatory bowel disease,  
 CC transplant rejection, nervous system disorders (e.g. Alzheimer's  
 CC disease), ischemic conditions, metabolic disorders (e.g. obesity and  
 CC diabetes) and infertility. The invention is useful in gene therapy. The  
 CC present sequence is rat C3b/C4b complement receptor like protein

XX Sequence 3095 AA:

Query Match 63.1%; Score 10720; DB 5; Length 3095;  
 Best Local Similarity 64.3%; Pred. No. 0;  
 Matches 1843; Conservative 436; Mismatches 543; Indels 44; Gaps 5;

QY 195 DGGGDDGQKTYLVMSONACSDSPHTPGRIPE---SSMGDWKRWYLTLCRISSDA 251  
 DB 1 DAGVGGDTRSLVLY-----LTGSSVPDLIVSMNOMWLH-----LQSDDS 40  
 QY 252 RSGSVRSKPKTSNVELVAPG-----TEIQGSCGDPGIPAYGRREGSRFHHGDTLRFEC 306  
 DB 41 -----IGSPGFAVYGEIEKGCGGDPGIPAYGKRTGSSFLHGDTLRFEC 84  
 QY 307 QPAELVGGKAITQKNNOMSAKRPQVSCFNFPTSGVLSPTNEDYGNHLACVWL 366  
 DB 85 QPAELVGERVITQKNNOMSGNKPSCVFCFNFPTASSGIIISPNTPEEGNMMNCVWL 144  
 QY 367 ILAPESRIHLAFNDIVEPOFDLVTKDGTAAAPVLTGTSNQQLSSITSGHVARLE 426  
 DB 145 IISPSGRIRHLIFNDVPEQDFDLAVKDDGISDITVLTGTSNEVAPQLXSHIVALE 204  
 QY 427 FQTHSTGKGFNTFTFRHNECPDPGVNPKGRFGDSLQLGSSISFLCDEGLTQGS 486  
 DB 205 FQSHSTGKGFNTFTFRHNECPDPGVNPKGRFGDSLQLGSSISFLCDEGLTQGS 264  
 QY 487 ETTTCVLKEGSAVNNASVLRCEAPCGHLTSPSGTILSPGPFYKDALSCAWYIEAQP 546  
 DB 265 ESTICIIQDGNVWVSTVPCREAPCGHLTSSGVIIIPGMPGYKXSLNCENYIEAKPG 324  
 QY 547 YPIITDRPKTEVNYVTLFVRDRTYSAPLIGVYHQVPOPLISNLYLTFSDKS 606  
 DB 325 HSITITDRPKTEVNYVTLFVRDRTYSAPLIGVYHQVPOPLISNLYLTFSDKS 384  
 QY 607 HSDIGFOLRYETITLQSDHCLDPGIPVNGQRHNDPVYALVTSCSGYTLSDGEPLC 666  
 DB 385 RASVGFILHESVTLSDSCLDPGIPVNGQRHNSNPGIRSVTSCDPTGLSDDDLIC 444  
 QY 667 EPNFQSRALPSCALCGGFIQSSSGTILSPGPDYPPNNLNCWTIETSHGKGVFTFH 726  
 DB 445 EKNQNMHALPSCALCGGFIHGSSTVILSPGPDYPPNNLNCWTIETSHGKGVQNMFH 504  
 QY 727 TFFHESGHDVLLITENGSFPOPLQLQTSRLPAPISAGLYNFTNQRFTISDEMSYEGF 786  
 DB 505 TFFHESGHDVLLITENGSFSEPPVARLTGSVLPHTIKGLFENFTAOIRFISDSFISYEGF 564  
 QY 787 NITSEVLDLEPCEPEVPAYGIRKLOFGVGDVLTFCFPGYRLEGTARITCLGGRRLW 846  
 DB 565 NITAEVLDLEPCEPEVPAYGIRKLOFGVGDVLTFCFPGYRLEGTARITCLGGRRLW 624  
 QY 847 SSPLRCAVACGNSVTGTGTLSPNFPVNNNNHCECTYSIQTPGKGIQIKARAFELSE 906  
 DB 625 SAPLRCAVACGNSVTGTGTLSPNFPVNNNNHCECTYSIQTPGKGIQIKARAFELSE 684  
 QY 907 GDVLYKVDGNNNSARLLGVFSHSEMGVTLNLTSSSLMDPITTAENTSKGFEHSSFE 966  
 DB 685 GDVLYKVDGNNNSARLLGVFSHSEMGVTLNLTSSSLMDPITTAENTSKGFEHSSFE 744  
 QY 967 LIKCEDGTGPKFGYKVDHGHFAGSSVSPCDPGYSIRGSEBELICTLSEGRRTWDRPTC 1026  
 DB 745 LYKCEDGTGPKFGYKVDHGHFAGSSVSPCDPGYSIRGSEBELICTLSEGRRTWDRPTC 804  
 QY 1027 VAECGGTVRGEVSGQVLSPGYRABYEHNLNLCIWTIEAAGCTIGLHFLVPTDEVDHLR 1086  
 DB 805 VAECGGTVRGEVSGQVLSPGYRABYEHNLNLCIWTIEAAGCTIGLHFLVPTDEVDHLR 864

QY 1087 IMPDVEGVLKELSGPALPKDLHSTFNSVLTQSTDFPFTSKOGFALIOFSVSTATSCND 1146  
 DB 865 VMDGPVDSNILLKEMSSALPEDIHSTFNSLTQFSDFTISGSGFISQISTSLASTCND 924  
 QY 1147 PGIPQNSRSGDSWEADSTVFOCDPGYALQSAEISCVKIEENRFWQSPPTCIACGG 1206  
 DB 925 PGMPQNGTRYGDSREPDITFOCDPGYALQSAEISCVKIEENRFWQSPPTCIACGG 984  
 QY 1207 DLTPGSCVILSPNPPYPPGKCDMKVTVSPDVILALVININLEBGFPLHIYGRDS 1266  
 DB 985 NLTPGPVAVILSPNPPYPPGKCDMKVTVSPDVILALVININLEBGFPLHIYGRDS 1044  
 QY 1267 LSPILGFSYGOUPGRLESSNSLFLAFRSDASVNAAGFIADTENRRESCFPGSTLXG 1326  
 DB 1045 NSPLIGFSYGOUPGRLESSNSLFLAFRSDASVNAAGFIADTENRRESCFPGSTLXG 1104  
 QY 1327 TRVGSDLKGSSTVYCHGGEVEGTSTLSCILGPDGKPMWNNRPVCTAPCGGQVYVSD 1386  
 DB 1105 TRIGTDFKLGSTVYQCDGSGYKIVDPSEICVTGADGKPSMDALPACQAPCGGQVYVSE 1164  
 QY 1387 GVVLSPNYPQNYTSGQICLYFVTVPKDYVFGQAFPHHTLNDVYEVHDSHSHSLSS 1446  
 DB 1165 GVVLSPNYPQNYTSGQICLYFVTVPKDYVFGQAFPHHTLNDVYEVHDSHSHSLSS 1224  
 QY 1447 LSGSHTEESLPLATSNQVLIKFSKGLAPARGFHFVYQAVPRTSATQSSVPEPRYKRL 1506  
 DB 1225 LSGSHTEESLPLATSNQVLIKFSKGLAPARGFHFVYQAVPRTSATQSSVPEPRYKRL 1284  
 QY 1507 GSDFSVGAIVRFECNSGYALQGSPEIECLFVPGALQMNVSAPTCVPPCCGNTLERRGTI 1566  
 DB 1285 GSEFSAGSIYRFPENPBYLLQGSFALRCQGVNPAALQMNVTISCVPPCGSNFQRRGTI 1344  
 QY 1567 LSPGFPEPYNLNLCWIKVYVPGAGIQIOVVSFVTEQWDSLEVPDQADNTVYMLGSFS 1626  
 DB 1345 LSPGFPEPYNLNLCWIKVYVPGAGIQIOVVSFVTEQWDSLEVPDQADNTVYMLGSFS 1404  
 QY 1627 GTTVPALNLTSTNOLYHAFVSDISVSAAGFHLKXKTVGLSCEPAVPSNGVTVGERYLV 1686  
 DB 1405 GTTVPALNLTSTNOLYHAFVSDISVSAAGFHLKXKTVGLSCEPAVPSNGVTVGERYLV 1464  
 QY 1687 NDVVSFQCEBPGYALQGHAIHISCPGTVRNRYPPPLICIAOCGTVBEMEGVILSPGPN 1746  
 DB 1465 NDVVSFQCEBPGYALQGHAIHISCPGTVRNRYPPPLICIAOCGTVBEMEGVILSPGPN 1524  
 QY 1747 YPSNMDCSKIALPVGEGAHITQFLNLTSTENNDYIETRNPYETSRMGFSSGELPSSL 1806  
 DB 1525 YPSNMDCSKIALPVGEGAHITQFLNLTSTENNDYIETRNPYETSRMGFSSGELPSSL 1584  
 QY 1807 LSTSHETTVFHSDDHSGNRPGFLEYOAYELOECRDEPPRANGIVGAGVNVGQSVFEC 1866  
 DB 1585 LSTSHETTVFHSDDHSGNRPGFLEYOAYELOECRDEPPRANGIVGAGVNVGQSVFEC 1644  
 QY 1867 LPGVLTGHFVLTQGHNTNRMDHPLKPCVPCGNTSNGTVSAGFPSPSSQDCV 1926  
 DB 1645 YPGVLTGHFVLTQGHNTNRMDHPLKPCVPCGNTSNGTVSAGFPSPSSQDCV 1704  
 QY 1927 WLITVPIGHGVNLNLSLQTPESGDFITTVDGQOQTAAPRLGVFTRMAKKTVOSSNOVL 1986  
 DB 1705 WLITVPIGHGVNLNLSLQTPESGDFITTVDGQOQTAAPRLGVFTRMAKKTVOSSNOVL 1764  
 QY 1987 LKPHRDAATGIGFALFSAFALPKPCPPPTLLPAAEVTEBENIGDIYVRYCIPGTV 2046  
 DB 1765 LKPHRDAATGIGFALFSAFALPKPCPPPTLLPAAEVTEBENIGDIYVRYCIPGTV 1824  
 QY 2047 GNEILTKLGTLYLQFEGPPTICEVHCTNELLDSTGVILSOSYPSGYPQOTCSMLVRY 2106  
 DB 1825 GNEILTKLGTLYLQFEGPPTICEVHCTNELLDSTGVILSOSYPSGYPQOTCSMLVRY 1884  
 QY 2107 EPDYINISLTYEYLSKQXDBEPIFDGSGQSLKALSGNYAPALITVSSNSVYLRWS 2166  
 DB 1885 EPDYINISLTYEYLSKQXDBEPIFDGSGQSLKALSGNYAPALITVSSNSVYLRWS 1944  
 QY 2167 SDHAYNRKGFKIYSAFYCSLPPAPLHGFILGDTSTQPGSSIHFGCAAGRYLVGHSALIC 2226

```

Db      1945 TDHATSKGFKIRVAPYCSLSITLTKNGVLNKTAGALGSKVQVFCRKGIMHSHNATC 2004
Qy      2227 TRHPGCHLMBEALPLCOALSCGLEPARKGMVFGEKEYTVTKAVYSCSEGYHLQAGAEA 2286
Db      2005 RRPVGVQWQSMAPLCOAVSCGIPKAPGNGSFTGNFTLDSKYTECNBEGFKLDAISOQA 2064
Qy      2287 TAECIDTGLMSNRNVPQCVPTCPDVSSISVEHGRRLIPEYQVQAOQLMLCDPGY 2346
Db      2065 TAVQOEDGLMNRKRPCTCKVPKPSIEGQISEHMLKLVGSLNEGAOVLSCSPGYF 2124
Qy      2347 YTGQVRIRCOANGKMSLSDSTPTCR11SCGELPIPNHRIIGILSVYATATFSCNGYT 2406
Db      2125 LOGQRLLQCCANGTMTNTEEDRPCKVISCGLSPFPNGKIKGLTLMGATATFTCNIGYT 2184
Qy      2407 LVGRVRBEMANGIMSGSEVRACLAGHGTPEPIYNGHINENSVYRSVYQCNAGPRLI 2466
Db      2185 LVSHVBECLANGIMSGSETRCLAGHCSGPDPIYNGHISGPGFSYRDTVYQCNPGFRLV 2244
Qy      2467 GMSVRIQQODHMGSKTPEVCVPTICGHPGNVNGLTQGNOPNLNDVYKFCVNGGYMAEGA 2526
Db      2245 GTSVRIQLQDHKMSGQTPVCPTICGHPGNHGLTNGSEPNLNDVNTCTHGYLQGA 2304
Qy      2527 ARSCCLASGQWMDLPTCR11INCTDPGHQENSVARQVHASGPHRSFGTTVSYRGNHGYL 2586
Db      2305 SRACRSNGQWSSPLPCR11VNCSDPGFVENAVHQQNFPESPEYGTSMYHCKKGFYL 2364
Qy      2587 LGTPVLSQQGDTWDRPPOCLIVSCGHPSPHSQMSGSLYVGNVYRSCIGKRLV 2646
Db      2365 LGSALTCMASGLMDRLPKLALSCGHPVANAVALTGLPTFGAVVQVSCGQQLTQ 2424
Qy      2647 NSTRMCGLDGHWTSGLPHCGSTSVGCGDGPAPAHGIRLSDSPGPTWAFSCGAGVLR 2706
Db      2425 NSTRVCOEDSHWSGSLPHCGSNPGFCDDPTPAHSGRLDDEKXTSLAFSCGEMHQRL 2484
Qy      2707 GSSRTTQANGSWGSGPECGVISCNPPTPSNARVVSGLVYFSSGIVYECGEGYATG 2766
Db      2485 GSARITLVGWSGVQVCEAVSCNPGPTNGMILSSGIIIFSSSVIYACWEGYKTS 2544
Qy      2767 LLSRHCSVNGTWTGSDDECLVINGDPIGIPANGRLANDPRNKTVTYQCVPMMSHR 2826
Db      2545 LMRHCANGTMTGTADCTIIISGDDGTLPNGIGRETDPTFNKIVSYQCNPGYLMPPRT 2604
Qy      2827 VSVLSTCKDRTWNGTKFVCKALMCKPPLIPNGKRVSGSDPMWSSVYACLEGYQLSLPA 2886
Db      2605 SPTRICKDGTMTQTRPLCKAVILCSOPPSVPGNKVSGSDPRMGASISYSCVDSGYQLSHSA 2664
Qy      2887 VFTCEGNGSWTGLPOCFPVFCGDPGVPSRGRREDGFSYRSVSFSCHPPLVLVGSPPR 2946
Db      2665 ILSCEGRGVKGVPOCLPVFCGDPGTPAGRLSGKSFYFKSVFTQCKRPPFLVGSRR 2724
Qy      2947 FCQSDGTWGTQSDICIDPTLTTCADBGVPQFGIQQNSQGVYGVTVLFCQKGYLLQGST 3006
Db      2725 TCQADGWSGIGTCTIDPAHTACPDGTPHFGIQQNSKGEVSGTYFFCRCKGYHIGST 2784
Qy      3007 TRTCLPNTLWSSGTPPOCVPHHCQPTPTHTANGALDPSMGYTLI 3052
Db      2785 TRTCLANLWTSIGITECTIPACRQPTPAHADVAIDLPAGTYTLV 2830

```

```

RESULT 15
AAE20787
ID AAE20787 standard; protein; 3069 AA.
AC AAE20787;
XX
XX
XX 01-JUL-2002 (first entry)
DE Human C3b/C4b complement receptor like protein #1.
KW Human; C3b/C4b complement receptor-like molecule; immune system disorder;
KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;
KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;

```

```

KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;
KW transplant rejection; autoimmune disease; ischaemic condition; neotrophic;
KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;
KW infertility; vasodilator; obesity; cardiac.
XX Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 387 /label= Unknown
FT Misc-difference 461 /label= Unknown
FT Misc-difference 586 /label= Unknown
FT Misc-difference 1272 /label= Unknown
FT Misc-difference 1272 /label= Unknown
XX MO200210199-A2.
XX
XX 07-FEB-2002.
XX
XX 24-JUL-2001; 2001WO-US023232.
XX
XX 02-AUG-2000; 2000US-0222504P.
XX 28-NOV-2000; 2000US-00728787.
XX
XX (AMGE-) AMGEN INC.
XX Welcher AA, Elliott GS,
XX N-PSDB; AAD33318.
XX WPI; 2002-303934/34.
XX DR N-PSDB; AAD33318.
XX
XX PT Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic
XX PT acid molecule, useful for treating, preventing and diagnosing rheumatoid
XX PT arthritis, psoriatic arthritis, inflammatory arthritis, and multiple
XX PT sclerosis.
XX
XX PS Claim 13; Fig 1; 251pp; English.
XX
XX CC The invention relates to a nucleic acid encoding a novel C3b/C4b
XX CC complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like
XX CC polypeptide and nucleic acid molecules may be used to treat, prevent,
XX CC ameliorate, diagnose and/or detect diseases such as immune system
XX CC disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory
XX CC arthritis, osteoarthritis, inflammatory joint disease, autoimmune
XX CC disease, multiple sclerosis, lupus, inflammatory bowel disease,
XX CC transplant rejection, nervous system disorders (e.g. Alzheimer's
XX CC disease), ischaemic conditions, metabolic disorders (e.g. obesity and
XX CC diabetes) and infertility. The invention is useful in gene therapy. The
XX CC present sequence is human C3b/C4b complement receptor like protein. Note:
XX CC ID NO:2 (AAE20900) shown in page 176-189 of the specification. However
XX CC the sequences differ at position 695
XX
XX SQ Sequence 3069 AA;

```

```

Query Match 63.0%; Score 10701.5; DB 5; Length 3069;
Best Local Similarity 65.7%; Pred. No. 0;
Matches 1833; Conservative 419; Mismatches 531; Indels 5; Gaps 1;
Qy 270 APG-----TELEQSGCGDPGIPAYGRREGSFRHHGDTLKEFCQAPAFELVQKAITCOKN 324
Db 17 SPGRFAYVQIEIKGCGCDPGIPAYGKRTGSSFLHGDLTITFCRPAFLVERVITCOQN 76
Qy 325 QMSAKKFCVNSCFPNTSPSGVTLSPNYPEDYGNHLHCVWLIIARDESRIHLAFNDIV 384
Db 77 QMSGNKPSCVFSCFPNTFASGIIISPNYDEYGNMNCVWLIISEPSRIHLIFNDIV 136
Qy 385 EPQDFLVIRKQGAFAEAPVIGTSPGNTLPSISGSHVAREQTHSTGRGNIFET 444
Db 137 EPQDFLVAVDDGDISDITVLETGTSNVEPQGLASSGHVIRLEFQSDHSTGRGNITTT 196

```

QY 445 FRHNECDPGVNVNKGKFGDLSLOLSSISFLCDEGFLTGOSENTITCVLKEGSVNMSAV 504  
 Db 197 FGQNECHDPGIPIGRBFGDRFLIGSSVFRHCDGFKTQGSSEITITCILDGNNVMSSTV 256  
 QY 505 LRCEAPCGHILTSBGTILSPBMBGFYKDALSCAMVIEAOPGYIKITFDRFKTEVNDT 564  
 Db 257 PRCAFPCCGHILTASGVILPFGMFGYKDSLHCMIIEAKGHSIKITPRFOGFEVNDT 316  
 QY 565 LEVRDATTYAPLIGVHGTQVPOPLISTSNVYLSTDSKSHDITFOLRYETITLQSD 624  
 Db 317 LEVRDAGASSPLIGBHYGTQAPQLISTGNFMVLTFTDNRSSISIFLHYEVTLESD 376  
 QY 625 HCLDPGIPVNGQRHGNFVYGALVTFSCDSCGYTSDDEPLECEBNFQMSRLPSCALCG 684  
 Db 377 SCIDPGIPVNXHNRHGGDFGRISTYTFSCDCEGYTSDDEPLVCENHOMHNLPSCDLGC 436  
 QY 685 GFIOGSSGTLISPGPDFFYPNNLNCWIIETSHQGVFTFHTPHLESBGHYLITENG 744  
 Db 437 GYIOGKGTVLSPGFDPFYPNSLXWTITVSHQGVOMI FHTPHLESSHDYLLITEDGS 496  
 QY 745 FTQPLRLQTSRLPAPISAGLYGNFTQVRFISDFSNGYSGFNITTEBYDLEPECEBPV 804  
 Db 497 FSBVVALTGSVLEPHITKAGLFGNFTQVRFISDFSNGYSGFNITTEBYDLEPECDGPV 556  
 QY 805 AYSRKGLQGVGDTLTFSCPGYRLBGTARITCLAGRRRLMSPLPRCAECGNSVTGT 864  
 Db 557 AFBRICFPHFGVDSLTFSCFLGRLBGAKKLTCLGGRRWMSAPLPRCAECGNSVTGT 616  
 QY 865 QGTLISNPNVNNNNECIYISIOTOPKGIOLKARAFELSEBGVLYKVDGNNNSARLLG 924  
 Db 617 EGTLSNPNPNVNNNNECIYIKIETEAQKGIHLATRSFOLFEQDTLKYVQKSSHPLG 676  
 QY 925 VFSHSEMMGTVLNSTSSLMIDFTDABNTSKGELHFSFELIKCEDPCTPKKGYVHD 984  
 Db 677 TFTNELLGLILNSTSHMLBFTNNGSDTDQGFOLTYTSPDLKCDPGIPNNGYIRD 736  
 QY 985 EGHFAGSSVSFSCPGYSLRGSEELLCLSGERTWDRPLPTCVNACGTYRAGEVSGVLS 1044  
 Db 737 EGHFTDVLVYSCNPGYAMGNSNTLCLSGDRWDRPLPSCLAECCGQIHAATSGILS 796  
 QY 1045 PGYAPAYEHLNLCMTIETBAEGCTIGLFLVFTDEVDVLRIMDGPVESGVLLKELSGP 1104  
 Db 797 PGYAPAYDNNLHCMTIETADPGKTI SLHFI VFTDEMAHDLIKWMDGPVDSOILLKEMSGS 856  
 QY 1105 ALPEDLHSTNSVTLQSTDFTFSTKQCPALIOFSVSTASCNDRPCTPONGSRGDSWAGD 1164  
 Db 857 ALPEDLHSTNSVTLQSTDFTFSTKQCPALIOFSVSTASCNDRPCTPONGSRGDSWAGD 916  
 QY 1165 STVFCQCPGYALQGSABISCVKIEENFFWOPSPPTCIAPCGGDLTGBSGVILSPNYEPY 1224  
 Db 917 TMTFCQCPGYALQGSABISCVKIEENFFWOPSPPTCIAPCGGDLTGBSGVILSPNYEPY 976  
 QY 1225 PRGKCDMKVTVSDVYIALVFNIEPEGYDLFIHYDGRDLSPLIGSPFGSOLPGRIE 1284  
 Db 977 PRGKCDMKVTVSDVYIALVFNIEPEGYDLFIHYDGRDLSPLIGSPFGSOLPGRIE 1036  
 QY 1285 SSSNSLFLAFRSDASVSNAGFVIDYENPRESCTDPGSIKXKTRVSGDLKLGSSVTVCH 1344  
 Db 1037 SSSNSLFLAFRSDASVSNAGFVIDYENPRESCTDPGSIKXKTRVSGDLKLGSSVTVCH 1096  
 QY 1345 GGYVEBGTSTLSCILGPGDKVYNNRPVCTAPCGGQVSGDGVLSPNYPONTSOIC 1404  
 Db 1097 GGYVEBGTSTLSCILGPGDKVYNNRPVCTAPCGGQVSGDGVLSPNYPONTSOIC 1156  
 QY 1405 LYFVTVKDYVVFQGFATLALNDVVEVHDGHSQHSRLLSLSGSHTGESPLATSNQV 1464  
 Db 1157 LYFVTVKDYVVFQGFATLALNDVVEVHDGHSQHSRLLSLSGSHTGESPLATSNQV 1216  
 QY 1465 LKRSAGGLAPARGFHYQAVPRTSATQCSVPEPRYGRKLGDSFVGAIVRECSNGY 1524  
 Db 1217 LKRSAGGLAPARGFHYQAVPRTSATQCSVPEPRYGRKLGDSFVGAIVRECSNGY 1276  
 QY 1525 ALQSPRIECLPVFGALAQNMVSAPTCVVPCGGNLTERRGITLSPGPEPYLNSLNCVWK 1584

Db 1277 LLQGSTALHQSPNMLAQNMNDITPSCVVPSCGNFTQGRGTLSPGVPEYGNLNCVWK 1336  
 QY 1585 IYVPEGAGIOIQVSVFTEBQNMDSLEVFDADNTVMLGSFSTTVPALINSTSNOLYLH 1644  
 Db 1337 IYVPEGAGIOIQVSVFTEBQNMDSLEVFDADNTVMLGSFSTTVPALINSTSNOLYLH 1396  
 QY 1645 FVSDISVSAAGPFLKXYTGLSCPEPAVSNQVTKGERLVNDVVSFOCEPCTALQGH 1704  
 Db 1397 FVSDISVSAAGPFLKXYTGLSCPEPAVSNQVTKGERLVNDVVSFOCEPCTALQGH 1456  
 QY 1705 HISCMGTARMMYPPPLCIAQCGTVEEMEGVILSPGPBNYPSNMDCSMKIALPVFG 1764  
 Db 1457 HISCMGTARMMYPPPLCIAQCGTVEEMEGVILSPGPBNYPSNMDCSMKIALPVFG 1516  
 QY 1765 AHIOPLNFSTEPNHDYIEIRNGYEFYSRMKGRSGSELPSLSLSTHETTVYHSDHSON 1824  
 Db 1517 AHIOPLNFSTEPNHDYIEIRNGYEFYSRMKGRSGSELPSLSLSTHETTVYHSDHSON 1576  
 QY 1825 RPFKLEYOAYELOCPDPEPFANGIVRGAVNVGGSYFECPLGYOLTGHPLVTOOHT 1884  
 Db 1577 RPFKLEYOAYELOCPDPEPFANGIVRGAVNVGGSYFECPLGYOLTGHPLVTOOHT 1636  
 QY 1885 NRAMDHPKCEVPCGANTSSNGTVYSPGFPSPYSSODCWMLITVPIGHVRLNLSL 1944  
 Db 1637 NRAMDHPKCEVPCGANTSSNGTVYSPGFPSPYSSODCWMLITVPIGHVRLNLSL 1696  
 QY 1945 QTEPSGDTITWDGPOQTAARLQVFTRSNAKKTVOSSSNOLIKFHRDAATGCI PAIABS 2004  
 Db 1697 QTEPSGDTITWDGPOQTAARLQVFTRSNAKKTVOSSSNOLIKFHRDAATGCI PAIABS 1756  
 QY 2005 AYPLTCRPPITLIPNNEVTENEEPNIGDIVRRCILPGFLVGNELITCLGTYLOECP 2064  
 Db 1757 AYPLTCRPPITLIPNNEVTENEEPNIGDIVRRCILPGFLVGNELITCLGTYLOECP 1816  
 QY 2065 PPICEVHACPTNELITSTGVILISQSYPGSPQOTCSMTLVRVBDNII SLTVEYFLSEKQ 2124  
 Db 1817 PPICEVHACPTNELITSTGVILISQSYPGSPQOTCSMTLVRVBDNII SLTVEYFLSEKQ 1876  
 QY 2125 YDEFEI FDGSGSPILKALSGNYSAPLYTSSNSVYLLRMSDDAHNRKGFIRISAPY 2184  
 Db 1877 YDEFEI FDGSGSPILKALSGNYSAPLYTSSNSVYLLRMSDDAHNRKGFIRISAPY 1936  
 QY 2185 CSLPRAHLGFIIGSTORPGSIHFGCNAQVYLVNHSMLICRHPQGVYLMSEALPLQ 2244  
 Db 1937 CSLPRAHLGFIIGSTORPGSIHFGCNAQVYLVNHSMLICRHPQGVYLMSEALPLQ 1996  
 QY 2245 ALSGCLPEAPKXGMVFEKETTGTAKVYSCSEGYHLQAGAEATACIDTGLMGNRPVQ 2304  
 Db 1997 ALSGCLPEAPKXGMVFEKETTGTAKVYSCSEGYHLQAGAEATACIDTGLMGNRPVQ 2056  
 QY 2305 CVBPTCEVSSISVEHGRWRLIETQYQFOALMLICDPGYTYTGORVIRCOANKWSLG 2364  
 Db 2057 CVBPTCEVSSISVEHGRWRLIETQYQFOALMLICDPGYTYTGORVIRCOANKWSLG 2116  
 QY 2365 DSTPTCRIISCGSLPIRPINHRIGTSLVYCATLIFGNSGYTLVGSRYVRECMANGLSGS 2424  
 Db 2117 DSTPTCRIISCGSLPIRPINHRIGTSLVYCATLIFGNSGYTLVGSRYVRECMANGLSGS 2176  
 QY 2425 EYRCLAGHCTPEPIVNGHINGENYSYRGSVVYOCNAGFLIGMSVRI COODHMSGKT 2484  
 Db 2177 EYRCLAGHCTPEPIVNGHINGENYSYRGSVVYOCNAGFLIGMSVRI COODHMSGKT 2236  
 QY 2485 FCVPITGHPNPNGLTQGNQNLNDVYKVCNPGYMAEGAARSOCLASGOWSDMLPTC 2544  
 Db 2237 FCVPITGHPNPNGLTQGNQNLNDVYKVCNPGYMAEGAARSOCLASGOWSDMLPTC 2296  
 QY 2545 RIINCDRPHQENSVQVHASGPHRPSFGTVSVYRCHNGHYLIGTVPLSCGGSGTMDRPR 2604  
 Db 2297 RIINCDRPHQENSVQVHASGPHRPSFGTVSVYRCHNGHYLIGTVPLSCGGSGTMDRPR 2356  
 QY 2605 POCLIVSCGHPGSPHSOMSGDSYTVGAVVRYSCIGKRTLVGNSTMCGLDGMWTSGLPH 2664

Db 2357 PKCLAI SCGHPV PANAVLTGELTFTYGA VVHYSCRGSESLIGNDRVQCQEDSHWSGALPH 2416

Qy 2665 CSGTSVGVCGDPGIPAHGIRLGDSPDDGTVMRFSCGAGHYLRGSSERTCOANGSWGSGOP 2724

Db 2417 CTGNMPPFCGDPGTPAHOSRLGDDFKTKSLRFSCEMHQLRGS PERTCLNGSWGSLQP 2476

Qy 2725 ECGVISCNPGTPSANARVFSDDLFFSSIVYECREGYATGLSRHCSVNGTWGSDPE 2784

Db 2477 VCEAVSCGNPGTPTNGMIVSSDGLFSSSVIYACMEGYKTGMLMTRHCTANGTWGTAPD 2536

Qy 2785 CLVINGDPRGIPANGRLANDFRNKTVTYQCVPGYMMESHRSVLSCTKORTWNGTKPV 2844

Db 2537 CTIISGDPGTGLANGIOFGTDFTFNKTVSYQCNPGYMEAVTSATIRCTKDGWNPSKPV 2596

Qy 2845 CKALMKRPPLIPNGKTVGSDPFMMGSSVTYACLEGYOLSLPAVPTCEGNGSWTGELOCF 2904

Db 2597 CKAVLCQPPPVQNGTVEGSDFRMGSSISYSCMDGYQLSHSALISCRGVMKGBLPQCL 2656

Qy 2905 PVFCGDPGVBSRGREDRGSYSRSSVAFSCHPPLVLVGSPPRFQSDGTWSTGTOPSCIDP 2964

Db 2657 PVFCGDPGIPAEGRLSGKSFYKSEVFFQCKSPILVGSRRVQADGTWSGIQPTCIDP 2716

Qy 2965 TLTTCADPGVPOREGIOMNSQYQVSTVLRCCQKGYLLOGSTRTRTCLPNLTWGTBPDV 3024

Db 2717 AHNTCPPDPGTPHEGIONSSRGYEGSTVFFRCRGRYHIOGSTTRTCLANLWMSGIOJTECI 2776

Qy 3025 PHHCROPEPTPHANVGALDLPSMGYTLI 3052

Db 2777 PHACROPETPAHADVRALDLPFTGYTLV 2804

Search completed: October 18, 2004, 10:21:45  
Job time : 125 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model1

Run on: October 18, 2004, 10:04:11 : Search time 35 Seconds

(without alignments)  
8533.052 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985

Sequence: 1 MAGAPFPALLPCLISDCC.....RSGPVGDPSTLPGRHSPKP 3104

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR.79:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description               |
|------------|--------|-------------|--------|-------|---------------------------|
| 1          | 1555   | 9.2         | 2489   | 2     | 173012 complement C3b/C4b |
| 2          | 1473   | 8.7         | 3623   | 2     | 109456 intrinsic factor-B |
| 3          | 1404.5 | 8.3         | 3623   | 2     | 108618 intrinsic factor-B |
| 4          | 1390   | 8.2         | 2014   | 2     | 136936 complement recepto |
| 5          | 906.5  | 5.3         | 1025   | 1     | A43526 complement C3d/Eps |
| 6          | 904    | 5.3         | 1091   | 1     | PI0009 complement C3d/Eps |
| 7          | 763.5  | 4.3         | 991    | 2     | 152657 complement C3d/Eps |
| 8          | 755    | 4.4         | 991    | 2     | 152657 complement C3d/Eps |
| 9          | 746    | 4.4         | 991    | 2     | 152657 complement C3d/Eps |
| 10         | 721    | 4.2         | 1231   | 1     | NBHUH complement factor   |
| 11         | 685.5  | 4.0         | 3871   | 2     | T22812 complement factor  |
| 12         | 677    | 4.0         | 676    | 2     | A45900 complement C3b rec |
| 13         | 676.5  | 4.0         | 1234   | 1     | NBMSH complement factor   |
| 14         | 657.5  | 3.8         | 830    | 2     | A50359 P-selectin precurs |
| 15         | 640.5  | 3.9         | 1070   | 2     | T31069 P-selectin precurs |
| 16         | 626.5  | 3.7         | 2083   | 2     | T42721 CRP-ductin-alpha P |
| 17         | 605    | 3.6         | 768    | 2     | A42755 P-selectin precurs |
| 18         | 597    | 3.5         | 768    | 2     | I53821 P-selectin - rat   |
| 19         | 577.5  | 3.4         | 1057   | 1     | A39288 dorsal-ventral pat |
| 20         | 576    | 3.4         | 1464   | 2     | S58954 development protei |
| 21         | 556    | 3.3         | 1797   | 2     | T21889 hypotethical prote |
| 22         | 547    | 3.2         | 1805   | 2     | T21888 hypotethical prote |
| 23         | 545    | 3.2         | 1524   | 2     | T30337 polypotein - Afri  |
| 24         | 542.5  | 3.2         | 1053   | 2     | S46199 probable complemen |
| 25         | 526    | 3.1         | 610    | 2     | A35046 E-selectin precurs |
| 26         | 514    | 3.0         | 767    | 2     | T30018 hypotethical prote |
| 27         | 505.5  | 3.0         | 597    | 2     | S53711 C4BP alpha chain P |
| 28         | 504    | 3.0         | 707    | 2     | JC2218 procollagen C-endo |
| 29         | 501.5  | 3.0         | 1827   | 2     | T34288 hypotethical prote |

#### ALIGNMENTS

RESULT 1  
173012  
complement C3b/C4b receptor, membrane-bound form precursor - human  
N:Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CR1); surfac  
N:Contains: complement C3b/C4b receptor, secreted form  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence, revision 24-Nov-1999 #text change 09-Jul-2004  
C:Accession: 173012; 156203; A47602; S03843; A28507; A24748; B24748; C24748  
R:VIK, D.P.; Wong, W.W.  
J: Immunol. 151, 6214-6224, 1993  
A:Title: Structure of the gene for the F allele of complement receptor type 1 and seque  
A:Reference number: 156203; MUID:94065175; PMID:8245463  
A:Accession: 173012  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-683; 'X', 685-1133; 'X', 1135-1471; 'X', 1473-2489 <VIK1>  
A:Cross-references: UNIPROT:PI17927; UNIPROT:Q16744; GB:LI7418; NID:G306678; PIDN:AA8606  
A:Accession: 156203  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-683; 'X', 685-894; 'A', 896-1000; 1451-1471; 'X', 1473-2489 <VIK2>  
A:Cross-references: GB:LI7418; NID:G306678; PIDN:AA860694.1; PID:G306680  
R:Mong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; W  
J: Exp. Med. 169, 847-863, 1989  
A:Title: Structure of the human CR1 gene. Molecular basis of the structural and quant  
A:Reference number: A47602; MUID:89176869; PMID:2564414  
A:Accession: A47602  
A:Molecule type: DNA  
A:Residues: 1-41 <MON>  
A:Cross-references: GB:X14893  
R:Hourcade, D.; Meisner, D.R.; Atkinson, J.P.; Holers, V.M.  
J: Exp. Med. 168, 1255-1270, 1988  
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b rec  
type 1.  
A:Reference number: S03291; MUID:89010527; PMID:2971757  
A:Accession: S03291  
A:Molecule type: mRNA  
A:Residues: 26-584 <HOU>  
A:Cross-references: EMBL:X14362; NID:G30197; PIDN:CAA2541.1; PID:G736240  
A:Experimental source: Clone CR1-4  
R:Klickstein, L.B.; Barrow, T.O.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.  
J: Exp. Med. 168, 1699-1717, 1988  
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b  
A:Reference number: S03843; MUID:89035992; PMID:2972794  
A:Accession: S03843  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-894; 'A', 896-1000; 1451-2064; 'T', 2066-2276; 'P', 2278-2299; 'H', 2301-2325; 'T',  
A:Cross-references: EMBL:Y00816; NID:G30185; PIDN:CAA68755.1; PID:G30186  
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.  
J: Exp. Med. 165, 1095-1112, 1987  
A:Title: Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domain

A:Reference number: A28507; MUID:87168191; PMID:2951479  
A:Accession: A28507  
A:Molecule type: mRNA  
A:Residues: 955-1221, 'FV', 1224-2064, 'I', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2  
A:Cross-references: GB:X05305; NID:930196; PIDN:CAA28593.1; PID:9809019  
R:Wong, W.W.; Klicstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.  
P:Title: Identification of a partial cDNA clone for the human receptor for complement fr  
A:Reference number: A94073; MUID:86067975; PMID:2933745  
A:Accession: A24748  
A:Molecule type: mRNA  
A:Residues: 311-333/729-745/831-845 <NO2>  
A:Cross-references: GB:M11569; NID:9180991; PIDN:AAA52297.1; PID:9180995; GB:M11617; NID  
C:Genetics:  
A:Gene: GDB:CR1; CD35  
A:Cross-references: GDB:119800; OMIM:120620  
A:Map position: 1q32-q33  
A:Interon: 411/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 6  
1464/2; 1513/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 2099/1  
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-2489/Product: complement C3b/C4b receptor, membrane-bound form #status predicted <M  
F:42-584/Product: complement C3b/C4b receptor, secreted form #status predicted <MATS>  
F:43-95/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:228-233/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-487/Domain: complement factor H repeat homology <FH07>  
F:493-549/Domain: complement factor H repeat homology <FH08>  
F:554-611/Domain: complement factor H repeat homology <FH09>  
F:616-682/Domain: complement factor H repeat homology <FH10>  
F:668-743/Domain: complement factor H repeat homology <FH11>  
F:747-803/Domain: complement factor H repeat homology <FH12>  
F:808-866/Domain: complement factor H repeat homology <FH13>  
F:871-937/Domain: complement factor H repeat homology <FH14>  
F:943-999/Domain: complement factor H repeat homology <FH15>  
F:1004-1061/Domain: complement factor H repeat homology <FH16>  
F:1066-1133/Domain: complement factor H repeat homology <FH17>  
F:1138-1193/Domain: complement factor H repeat homology <FH18>  
F:1197-1253/Domain: complement factor H repeat homology <FH19>  
F:1258-1316/Domain: complement factor H repeat homology <FH20>  
F:1321-1387/Domain: complement factor H repeat homology <FH21>  
F:1393-1449/Domain: complement factor H repeat homology <FH22>  
F:1454-1511/Domain: complement factor H repeat homology <FH23>  
F:1516-1582/Domain: complement factor H repeat homology <FH24>  
F:1588-1643/Domain: complement factor H repeat homology <FH25>  
F:1647-1703/Domain: complement factor H repeat homology <FH26>  
F:1708-1766/Domain: complement factor H repeat homology <FH27>  
F:1771-1837/Domain: complement factor H repeat homology <FH28>  
F:1846-1902/Domain: complement factor H repeat homology <FH29>  
F:1907-1964/Domain: complement factor H repeat homology <FH30>  
F:1969-2035/Domain: complement factor H repeat homology <FH31>  
F:2041-2096/Domain: complement factor H repeat homology <FH32>  
F:2100-2156/Domain: complement factor H repeat homology <FH33>  
F:2161-2219/Domain: complement factor H repeat homology <FH34>  
F:2224-2300/Domain: complement factor H repeat homology <FH35>  
F:2298-2354/Domain: complement factor H repeat homology <FH36>  
F:2359-2415/Domain: complement factor H repeat homology <FH37>

Query Match 9.2%; Score 1555; DB 2; Length 2489;  
Best Local Similarity 20.9%; Pred. No. 1.6e-79;  
Matches 679; Conservative 332; Mismatches 1021; Indels 1224; Gaps 142;

QY 58 KYSVLQYGVSGQHNNCPDGPPIPERGK-RLGSDFRIGSVQFPCNMGYLOQSKRITCKK 116  
DB KNSVWGAKDRCKRCKRNPPDPVNGVHVHVKIGIQGSIKYSCTKYRLIGSSSATCT- 146  
QY 117 VSDMFAWSDHREVCARMCDAHLRGPSSGILTSPPPIQYDNNNAHCWVIT-ALNPSKYI 175  
DB 147 ISGDVTYIMNETHICORIC-----GLPPTINGDPISTNRNFHSGVVTYKCPGSGG 201

QY 176 KLAPEFDELERGVDITLVGDGGDDQKTVLYNSQACSDSPHTPSRIPESSNGDIWQ 235  
DB 202 KRYFE-----LVGEPS-----ICTSN-----DQVG-IW-- 225  
QY 236 KMTVLEICRDISSDARSYSVKSPKTSNAVELVAPGTIEQSSGDP-----GIYAYGR 291  
DB 226 -----SGAPQCIIT-----NKCTPPVENGLVSDNR 253  
QY 292 EGSRFHNSDITLKECPAFELVGQKAITQKNNSAKKPGCVFSGFNFSPSGVLSAP 351  
DB 254 --SLFSLNVEVEPRCPGPFVMKPPRVKQALNKVEPELPSCSRVC-----QPPDVLHAE 307  
QY 352 NPEDEQNHLCWMLLAPESRIHAPNDIVPEQFDPVLKDGATAPALGTFSGNQ 411  
DB 308 RTQRDKNF-----SPGEVYS-----CEPGYD-----LRGAASMRCTPQGMS-- 347  
QY 412 LPSSTTSGHVARLEFQDHSSTGRKGFNITFTFRHNECPD-RGVPVNGK-RFGDSLQIG 469  
DB 348 -PAA-----PTCEVKSDDPMGOLLNQRVLFPVNLQIG 379  
QY 470 SSIPLCDGFLGTQGSSETITCVLKEGSVWNSAVLRCEAPCGHLTSPSGTILSPGWP 529  
DB 380 AKVDFVCEBGF-QLKSSASAYCVLAGMESLMSVPCV-----QIFGSPVPIWG-- 430  
QY 530 FYKDALSCMWLEAQPVPKIKITPRFKTEVYD-----TLVYR 568  
DB 431 -----RHTGKPLEVF--PFGKAVNYTCDPPDRDGTSPDLIGESTRICTSPQ 475  
QY 569 DGRYVS--APLIGVHGTQVPOFLISTNSNYLLPSTKSHSD-IGFOLRYE----- 617  
DB 476 GNGWSSPAPRCGILGHQAP-----DHLPLAKTKQTMSDFIGISLKECPREYVG 529  
QY 618 ---TTLGSDHLD-----PGIPVNGQRHG-NDYVGAIVTFSCDSG 655  
DB 530 RPSFIT-----CLDWLWSSPPDVCKRKSCKTPPDPVNGMVHITDIOGSRINYSCTTG 584  
QY 656 YTLISGEPLECEPN---PQWSRALPSCBAL-CGGFIQSSGTTILSPGPPDPYNNLNTCW 711  
DB 585 HHLIGHSABECLISGNAHMTKPTICGRICG----- 617  
QY 712 ILETSHGKGVFFPTFTFLHESGHDVLLITENGSPLOLUTGSRLPAPISAGLYGNFTA 771  
DB 618 -----LPTIANG----- 625  
QY 772 QVRFISDE-SMSYEGFNITPSSYDLPECEBEVPAYSIRKGLQREVGTLLPSCPPG-- 827  
DB 626 ---DPISTNRENPH-----YGSVVTYRCNPGSGG 651  
QY 828 ---YRLGSTARITCLGGR--LWSSPLPRCAEGNSVTGTQGLLSPN--FPVNYNN 880  
DB 652 RKVPELVPEBPIYCTSNDDQIGMISGPAQCIT-----PNKCTPPVWZNG 696  
QY 881 HECTIYSIOTOPKGIQLKARAFELSEGDLKVYDGNNSARLLGVFSHEMKVTLNSTS 940  
DB 697 -----ILVSDNRS-----LFSINEV----- 711  
QY 941 GSIWLPDITDAENTSKGFELHFFSEFLICKEDPQTPKFGYVHDEGHPAGSSVSPSCDPG 1000  
DB 712 -----VEFRCPG 719  
QY 1001 YSLNGSEELTCLSGERTWRPLPTCVAECGTVRGEVSGQVLSPPYAPAEHNLNCTWT 1060  
DB 720 FVMGPPRRVKQALNK--WEPELPSCGRVC-----QPP----- 751  
QY 1061 IEADAGCTIGHLVPDTEEVHDLRIWDGVEGVLKELSGPALPKDLHSTFNSVVLQ 1120  
DB 752 -----DVLIH----- 755  
QY 1121 FSTDFPFSKGFALQFVSSTATSCNDPCIGIPONGSRSDSWEAGDSTVPQCDPGALOGSA 1180  
DB 756 -----AERTQRDKDNFSPQGEVEVFSCEPGVGLRGA 786

QY 1181 EISCKIENREFFMOPSPPTCIAPCGGDLTGP--SGVILSPNYPEPPYPCKECDWKVTSP 1238  
 Db 787 SMRTCPGD---NSPAPTCVKSCHDDPMGOLLNKRVLFP----- 823  
 QY 1239 DYVIALVFNILNLEPGYDFLHIYDGRDLSPLIGSFYGSQUPGRIESSNSLFLA----- 1293  
 Db 824 -----VNLQAKVDFVCD-----EGFOLKG---SSASVCVLAGEMSL 858  
 QY 1294 PRSDASVSMAGFVLDYIENPRESCFDPGSIKNGTRVSDTL---GSSVTYTHG----- 1344  
 Db 859 WNSVSPVCEQIF-----CPSPVINGRHTGKPLEVPFGKTVNTYTCDPHPDRG 907  
 QY 1345 GGYVEGTSTLSCILGDPGKPYMNNPRVCTAPCGGQYVSDGVVLSPNYQNTSQIC 1404  
 Db 908 TSPILIGESTIRCHSDPDGNGWSSPAPRC-----GIL----- 940  
 QY 1405 LYFVTPKDYVVFQOFAFHATALNDVVEVDHGSQ--HSRLLSLSGSHGESLPLATSN 1462  
 Db 941 -----GHCOAPDHFPLFAKLKTQTNASDPFGTS- 968  
 QY 1463 QVILKFAKGLAPARGF-----HFVYQAVPR--TSAQCSSVPEPRYK-RLGSDPSVG 1513  
 Db 969 ---LKYECPREYGRPFSTICTLDNLWSS--PKDVCKRKSCKTPDPVNGMVHTITDIOVG 1024  
 QY 1514 AIVAFECNSGYALQSPREIECLPVGALADONVSAPTCV--VPCGANLTERGTILSPCF 1572  
 Db 1025 SRIVYSCCTGHRILGHSAECI--LSGNAAMHSTKPIQRIPLCPPIIANGDITSTRE 1093  
 QY 1573 EPLYNSLNCWKIIVPEBAGIIOIVVSFVTEBONNDSLEVPDADNTVMTLGSFSGTTPVA 1632  
 Db 1084 NFHYGS-----VVTYRCNPGSGRKVFE-----LVGEPS----- 1112  
 QY 1633 LUNTSQULHYFSDISVSAAHLEKYTGGLSSCEPR--AVPS-----NGYKTGE 1682  
 Db 1113 -IYCTSD-----DOYIWSGAPPOCIIPNKCTPPVNGEGLIWD 1151  
 QY 1683 R---YLVDVVSFOCEPGYALQGHANHISCMGTVRMNYRPPLCIAOCGGTVEEMEGYL 1739  
 Db 1152 NRSILFSLNVEVEFCQPGFVVKGRPRVYCQ--ALNKMEPLPSCSRVC----- 1197  
 QY 1740 SPGEFPGNYSNMDSWKIALFVFGAMIOPLNFSTEPNHDIETIRNGEYETSRMNGRPSG 1799  
 Db 1198 -----QPPPDVL----- 1204  
 QY 1800 SELPSSLLSTSHETTVYTHSDHSQNRPGFKLEYQAYELQECRDEPRPANGIVRAGATVNG 1859  
 Db 1205 -----HAERTOR-----DKDNF-----SPG 1219  
 QY 1860 QSVTFECLPGYQLGHPVLTCQHGNNRMWHDRLPKCEVPCCGNTSS--NGTVYSPGPPS 1917  
 Db 1220 QEVYYSCEPGYDLRGASMRCT--POGDMSPAFTCEVKSCHDDPMGOLLNKRVLFP--- 1273  
 QY 1918 PYSSSODCVMLITVPIGHVNLNLSLQTEBSCFETITMDGPOQTAPRLGFTSRMAKT 1977  
 Db 1274 -----VNLQAKV-----DFVCD--EGFOLKG----- 1294  
 QY 1978 VOSSNOVLKFNHDAATGIPATAFSAVPLTK---CPPTILLPNAEVTNEN--EENIGD 2033  
 Db 1295 ---SSASVCVL-----AGMESLWNSSVPVCEQIFCPSPVINGRHTGKPLEVPFGPK 1344  
 QY 2034 IVRFRCLP-----GFTLVGNEILTKL-----GTYLQFEGBPPICEV--HCPINELLTD 2080  
 Db 1345 AVNTTCDBHPDRGTSFDLIGESTIRCHSDPDGNGW---SSPAPRCGLIGHCOA----- 1395  
 QY 2081 STGVILSQYPSGYPOFQTCSMVLVRBEDYNISLTVETFLSEKQYDEFEIFDGSQGPL 2140  
 Db 1396 -----PD-----HFLPAKXTQTNADPPIGSLK 1420  
 QY 2141 LKALSUNVASPLITYSSNSYVLKWS--DHAYNKGFIRISAPYCSIPRALHGFILGQ 2199  
 Db 1421 YECPEEYVYGRPFSTICTDNLV---WSSPKDVCCKRS-----CKTPRPVNGMVHVI 1468  
 QY 2200 TSTQPGSIHFGCUNAGYRLVGHSMALCTRHPOGYNLMEALPLGQALSGCLPBAPKGMV 2259

Db 1469 TDIOVGRINVSCTTGHRILGHSAECILSGNTHA--WSTKPIQRIPLCGLPPIIANGDF 1527  
 QY 2260 FG---KEYTVGTAUVSCSEGYH---LQAGAEIACEL-----DTGLMS-----NR 2299  
 Db 1528 ISTRNENFHYGSVVTYACNLGSRGRKRVFELVGEPSIYCTSDNDQVIGMSPAPOCIIIPNK 1587  
 QY 2300 NVPP-----QCV-----VTC-----PDVSSIS--- 2317  
 Db 1588 CTPPVENNGILVSDNRSLSFLNVEVEFCQPGFVVKGRPRVYKCOALNKMEBELPSCSRVC 1647  
 QY 2318 -----VEHGRWRLIFETOYFOAQLMLICDPGYUYYGQRYIRCOANGKWSLGDSTPCR1 2372  
 Db 1648 QPPEILHGHETBHQDNFSPGOEVFVSCFEGYDLRGAALHCTPOGDMW--PEAPRCVAV 1705  
 QY 2373 ISC---GELPIPNHGRITGLSV--YATAIIFSCNSGYTLVGSRYRECMAANG--LWSGS 2424  
 Db 1706 KSCDDFLGOL---PHGVLPFLNLQAKVSPVDEBGFRLKSSVSHCVLVGMRSLWNS 1762  
 QY 2425 EVRCLAGHCTPPIVNGHNGE---NYSYRGSVYVOCN-----AGFRLIGMSVRIQOQ 2475  
 Db 1763 VPVCEHIFCPNPAILNGRHTGTPSGDIPLYKEIYTCDDHPDRGMTFNLIGESTICTS 1822  
 QY 2476 DHH-----MSGKTPFC-VPITCGH-----PGNPVNGLTQGNFNLNDVVKFVGNP 2519  
 Db 1823 DPHNGVWSSPAPRCLESLVAAGHCKTEQPPFASPTIPIINDF---EPVGTSLNVEGRP 1878  
 QY 2520 GYMAEGARSOCCLASGOWSMPLPTCIIINCTDPHGENSVROVHASGPHRFSTTYVSYR 2579  
 Db 1879 GYFGKMFPSIS--CLENLWSSVEEDNCRKRSCKPPEPENGWHTITD-----TOFESTVANS 1933  
 QY 2580 CNHGFYLLGTPIVYSCODG---TWDRPRPOCLLVSCGHPSPPHMSQMSGDSYT----- 2629  
 Db 1934 CNBGFRLIGSBSTTCLVSGNNVTWDKAPICEIISIC---BPPTISNGDYTSNNRISFH 1989  
 QY 2630 VGAVVYRSC-----IGKRLVGNSTRMCGLDGHWITGSLPHCSGTSVGCGDPG 2677  
 Db 1990 NGTVVITYQCHTGRPDGBOFLFELVGRSLEY--CTSDDDQVGVWSSPPRCIGSTN--KCTAPE 2045  
 QY 2678 IPAHGIRLDS---FDPGTWRRSCENAGHYLRGSSERTCOANGWSGSOPECGVISCNP 2734  
 Db 2046 V--ENAIIRVPGNRSEFSLTEIYVFRCPQPFVWGSHTVQCOYTNMGWGPXKLPHCSRV--COPR 2103  
 QY 2735 GTPENARVFS--DGLVFSSISYVECEBEGYVATGSLSRHCSVNGTWTGSDPECLVINCQD 2792  
 Db 2104 PEILHGHETLSHOUNFSPGOEVFVSCPSYDLRKAALHCTPOGDMWSEAPRCVKSCHDD 2163  
 QY 2793 --PGIPANGRLAGNDFYNNKTVITYQCVPGYMMESHVVS--VLSTCKDRTNNGTPEYCKAL 2848  
 Db 2164 FLQGLPHGRVULPLNLQAKVSVFCDBGFRILKGRASHCYLAQMK--ALMNSSVPPVCEQI 2222  
 QY 2849 MCKRPPLIPNGKVVGS---DFMGGSVTVAC-----LEGYLSLPAVTECGNG 2894  
 Db 2223 FCPNPAILINGRHTGTPFGDIPLYKEIISYACDTHPRGMTFNLILG--ESSIRCTSDROGNG 2281  
 QY 2895 SWTCELPOC---FVPEFGDGCVPBGRGRBERGRFSY---RSVSFSCHPRLVTVSPRRFC 2948  
 Db 2282 VMSSPAPRCLESLVAACPDPPKIQNGHYIGCHVSLVYPMWTISTICDPGYLVKGKFTFC 2341  
 QY 2949 QSDGTWGTGPOSCIDPLTTLTCADPGVPOFGIIONN---SOGYOVGSYVLPFCQKQYLLQGS 3005  
 Db 2342 TDGQIWSQDLHYCKE---VNCSPF--LPMNGISKELEMKKYYHGDYTLACEDQYTLBGS 2397  
 QY 3006 TTRTCLPNLTWSGTP 3021  
 Db 2398 PWSQCADDRWD--FP 2411

RESULT 2  
 109456  
 Intrinsic factor-B12 receptor Cubilin precursor - human  
 C/Species: Homo sapiens (man)  
 C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004

C:Accession: T09456  
 R:Kozyrak, R.; Kristiansen, M.; Silahcaroglu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N  
 Blood 91, 3593-3600, 1998  
 A:Title: The human intrinsic factor-vitamin B12 receptor, cubillin: Molecular characteriz  
 ion.  
 A:Reference number: 216677; MUID:98241400; PMID:9572993  
 A:Accession: T09456  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3623 <KOZ>  
 A:Cross-references: UNIPROT:O60494; EMBL:AF034611; NID:G3929528; PIDN:AAC82612.1; PID:G3  
 C:Genetics:  
 A:Map position: 10p12  
 C:Superfamily: intrinsic factor-B12 receptor cubillin; EGF homology  
 C:Keywords: receptor; vitamin B12 uptake  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-3623/Product: intrinsic factor-B12 receptor #status predicted <MAT>  
 F:436-467/Domain: EGF homology <EGF>

Query Match 8.7%; Score 1473; DB 2; Length 3623;  
 Best Local Similarity 21.2%; Pred. No. 1,2e-74;  
 Matches 795; Conservative 416; Mismatches 1184; Indels 1356; Gaps 174;

QY 113 TCMKSDNFPA-----WSDHRPVCA--RMCDAHLRGPSSGIIITSPNPPIQYDNNAHCWVI 165  
 DB 446 TCVGDVDSFSCBCECTRLMTG--ALCVPOQVCGESLGSINGSFYSRSPDVGVHDVNCFW 503  
 QY 166 ITANPSKYIKLAFEEPLER---GVDTLTVGCGGQDQKTYLVMSQNACSSEPTPG 221  
 DB 504 IKT-EMGVNLRITTFPLESMNDCPHEFLQV---YDGSSSAFQGRCCSLPH--- 555  
 QY 222 SRPESMGDIWROKMTVLEICRDISSDARSQSVKSPKTSNAVELAVGTETEGSSCG 281  
 DB 556 -----ELSSD-----NALYHLXSEHLRNG--- 576  
 QY 282 DPGIPAYGRGRSGRFHGDITLKEFCQPAFELVGKAITCQKNOWSAKKFCVPSCEPNE 341  
 DB 577 -----RGFTVR-----METQCP-----ECGILL 594  
 QY 342 TSPGVVLSPNYPEDYGNIHLHCWVLLARPESRIHLAFNDIVER-----QFDFLVIKDGA 397  
 DB 595 TGPFGSIKSPXPENYPRGDCWIVTSPDLVLTFTFGTSLSEHDDCKXDLLEIRDP 654  
 QY 398 TAEAPVLTSGNOLBSITSSGVNARLEPQTHSTGKGFNTFTTFRN--ECPPPGV 456  
 DB 655 LYQPLLKGFCTTFVSVPPLQTGPFARIHPSDSQISDQGFHTIYLTSPDLACGQNYTD 714  
 QY 457 VNGKRF-----GDSLQLG--SSISFLCD-----EGF 480  
 DB 715 PEGELFLPELSGPTHTRCQVVMKKOPQGEQIQINFTHEHLQCSOSSQNYIEVRDGETL 774  
 QY 481 LG-TQGETTTCVLRKGSVW-----NSAVLRC--EAPCGHLLTSPGIIISPGMP 528  
 DB 775 LGRVCGNGTISHIKSITNSVWIRFKIDASVEKASFAVVQVACGDLTG--EGVIRSPFP 833  
 QY 529 GFYKDALSCMWLEADRGVRIKTPDRK---TEVNYDTLEVROGRYTSAPLIGVYHGT 584  
 DB 834 NVYRGERTGKWTIHQPOSVILNFTVFEIGSSAHCEYVEIGSSSILGSEPNKKYCGT 893  
 QY 585 QVBPQLISTSNVLYLLFSTDKSHSDIGFOLRYETITLQSDHCDPPIPVNGQRHNDP 644  
 DB 894 DIFSFTSVYVNFLLYFVFKSSSTENHGFMAKF----- 925  
 QY 645 GALVTSQDSGYLLTSGEPLCEPFWQSRALPSCAL--CGGFIQSSGIIISPGPPFY 703  
 DB 926 -----SADDLACGEIILTESTGIIISPGHPNY 952  
 QY 704 PNNLNCWTIIETSHGKGVFTTPTHE-----SGHYYLITENGSTFQRLRLTSSKRLA 759  
 DB 953 PHQINCTWIIIVQPNHILHMETFTHEBTHVNCNDYIEVYDTDETS--LGYYCCSKSIP 1011  
 QY 760 PISAGLYGFTAQVRIIDFSMSYEGFNITFSEYDLEPCEPEVPAVYSIRKLGQGVGDT 819

DB 1012 SLTSS--GN-SLMLVPTVSDLAEGFLINYEAI----- 1042  
 QY 820 LTFSCFPGYLRBGTARITCLGGRRLMSSPLPRCAVECGNSVYTCGTLSPFPVYNN 879  
 DB 1043 -----SAATACI-----QDYTDIGTISNFPNNYNN 1070  
 QY 880 NHECIYSIQTPGKGIOLARAFELSEG-----DVLKYVDGNNNSARLLGVFSSHEMNG 933  
 DB 1071 NMECIYRITVTRTQOLIAVHFTNFSLEAIGNYYTDLERDGYEKSPLIGIF--YQSNLP 1129  
 QY 934 VTLNRSLSLWDFITDAINTSKGPELHSSBELIKCEDPGRPKRGYKVADBGHFAAGSV 993  
 DB 1130 PIIISNSKMLKFSQDQIDTSSGSAVWD-----GSSYGCQGNLTSSG 1174  
 QY 994 SFSCDPGY-----SLRSEELL-----CLSGERRTMDRP--- 1022  
 DB 1175 TF-ISPNYPMPIYHSECEYMWLKSHGSAFELFPDFLHEHNPCTLDVLAAYDDSSNS 1233  
 QY 1023 -----LPTCAVECGTV-----RG-----EVSQVLS 1044  
 DB 1234 HLLTQLCGDEKPLLRSSGDSWFIKLRDDEGQGGFKAERYQTCENNYIVNQTYGILES 1293  
 QY 1045 PGYAPYEHNLNCITIEAEAGCTIGLHFLVDEEVH-----DVLRIWDGVESGVLLK 1099  
 DB 1294 IGYPNYSINQCHMTIARTNTNTNYTFLEADLEH--HINSTDYLELYDGPROMG--- 1348  
 QY 1100 ELSGPALPKDLHSTNSVVLQFTDFTSKQ--GFAIQPSV-----STAT--SCNDPGIP 1150  
 DB 1349 RYCGVDLPPPGSTSSKLVLLLTGQVGRREKFGQMFVYVGGCELSGATGSSPSPFP 1408  
 QY 1151 QNGSRSGDS--WAGGSTYFQCDPGVALO-----GSAETSCVKI 1187  
 DB 1409 NRPYKKECIW-----YIRDPGSSIQTLTHDPDVEYHRCNPDVLEIYGGPPHSPRI 1462  
 QY 1188 -----EN-----RF-----FQPSPPTCIAPCGGLTSPS 1212  
 DB 1463 AQLCTGRSPENDPMQVSSYGNELAIRFKTDSLINGRGNFASQ---AVTGCGGIGFOAPS 1518  
 QY 1213 GYILSPNTPPEYRPEKEDDMKYTVSPDYVALVFNIFNLEPGVFLHIYDGRS--LSPLI 1271  
 DB 1519 GEIHSFNYSPEYRSNTDCSWAIRYDRYHRLVNLNFTDPLEPQDCIMAYDGLSSTMSRLA 1578  
 QY 1272 GSFYSSQLPGRLESSNSGLFLAFPSDASVSAGVIDVTEPRRSC-----FD--- 1319  
 DB 1579 RTCGEQLANIVSSGNSLFRFOSGRQNRGRAPF---RACGGHILITSSFDTVSS 1634  
 QY 1320 ---PESINRGR---VGSIDLKSSVYTYCHGGEVEGTSLSC---IL--GPDGKPV 1366  
 DB 1635 PRFPANYPNNQNCWIIQACPRLNHTLSFTH--FEIERSYTCARDVVELIDGHEAPL 1692  
 QY 1367 -----MNNRPVCT-----APCGQVYGSQVVLSP 1392  
 DB 1693 RGRVCGTDMPHRITSSSALTFLRFVSDSSISAGGFHTTVASVACGTFPMAGIINSP 1752  
 QY 1393 NYPNYTSQOICLY--FTVPEKDYVFGQFAFFHTAL-----NDVVEVHGSHQSRLL 1444  
 DB 1753 GYPDIVPNNVECWNIISPGNRL---QLSFIQLEDSODCSGDFEIRGNNA-----T 1804  
 QY 1445 SLSGSHTEBSLPLATNOV---LIFSAKGLAPARGFHYQAV----- 1486  
 DB 1805 GHLVGRYCGNSFPLNYSIVGHTLWAFIISDGSGGTGQATFMKIFGNDNIIVTHQKVA 1864  
 QY 1487 -----PRISAQO-----CSSVPEPRY----- 1502  
 DB 1865 SPFWPENYPIINSYQMTVNNVNNASHVHGRILENDIEIQNCYYDKRIYDGPISHARLIG 1924  
 QY 1503 ---GKRLGDSFYGAIVRECNAGVALQSPRETECLVPBQALAQW--NVSAPTCVVP--- 1554  
 DB 1925 AYCGTQTESPSSGNSLTFHYSDSSISGK-----GFLEWPAVADAPDVLTTIAP 1975  
 QY 1555 ---CGQNLTERGR---ILSPGPPEYLLNSKNCVWKIVPEAGAGIIOVVSFTEQN--- 1605  
 DB 1976 GACGGFL--RTGAPVFLRSPGWDYSNRVDCITWILQAPDST--VELNITLSDIESHRTC 2032

QY 1606 -MDSLVEVDGADNTVTMLGSPSGTTPALNLTSTNOULHFEYSDI SVSAGFHELYKTVG 1664  
 DB 2033 AYDYLVRDGDNNLAAQLAULCGREIPRISTGEBYFIRFTSSSVTRAFASFSH-- 2089  
 QY 1665 LSSCPREPAPNSGVKTERYLVDNVSPQC-----EPGYALQGH-----AHISCM 1709  
 DB 2090 -KSGCGYLHMDRGLITSPKPEYTPSNLNCMHWLVGSLTIAVHFEDPFIPIRGDSSCN 2148  
 QY 1710 PG---TYRRMN--YPPPL-----CIAO----- 1726  
 DB 2149 QGDYLVRNGBDIYSPPLGPPGNGHFCGSHASSLTFTSDNQMFVOFISDHSNEGQFKI 2208  
 QY 1727 -----CGSTV-----EEMEGVILSPGPGNYPNSNDCSKIALPFGFHNIF--LN 1771  
 DB 2209 KYEAKSLACGGNNTYIHDASAGYVTSRNHPNYPHPADCIWILAPETRIQLQFEDRFD 2268  
 QY 1772 FSTEPN--HDYIEIRNGPEYETSRMMGRPSGSELSLSTSHETTVVFHSDHSONRPGFK 1829  
 DB 2269 IEVTPNCTSNVLELRDGVDSAPILSKFCGTSLPSSQMSGGEVWYLRFRSDNSFTYHFK 2328  
 QY 1830 LEVAYAYELQECBPDEPPFANGIVRAGY-----NGQSVTFECLPGYQLT----- 1873  
 DB 2329 AK---YSIAQCGGVPGSGGVESIGHPTLPYRDNLFCEWHLOGLSGHYLTISFEDFNLO 2385  
 QY 1874 ---GHVPLTQGHGNNRMD-----HBLP----- 1893  
 DB 2386 NSSG---CEKDFVEIWDNHTNSGILGRYCGNTIPSIDTSSNTAVRFTVDSGVTASG 2440  
 QY 1894 -----KCEVPCGNITSSNGTVYSPGPPSYSSODCVMLITVPIGHVRL--NLSLQ 1945  
 DB 2441 FRLAFESMECCGDLQSGITFTSPYRNPNPHGRICEMRITAPBGRRIITLPMNRLA 2500  
 QY 1946 TEPS--GDFITWDGPOQTAPRLGVFTKSM--AKTVOSSNOVLKPHRDAAT--GGIFA 2000  
 DB 2501 THPCNNEHVIVFNGIRSNBPQLEKCSVAVVSNIEKSSGNTMKVIFFTDGRPYG--FT 2559  
 QY 2001 IAFAYAYELTCCPPTIILPNAVVTENEFNIGDIVRRCLEPFLNCG----- 2048  
 DB 2560 ASYTSSEDAVCGGS--LPN---TPEGNFT-----SPGYGVNRYSNLNCENTLS 2604  
 QY 2049 -----EILTCKLGYLOFEGP-----PRICE 2069  
 DB 2605 NPNQGNSSISIHFEFDYFLESHQDFVLEFRVC--DADGPELMMRLCGSKPTLPLVIP 2661  
 QY 2070 -----VHCPTNELI-----TD-----STGVILSQSYPGSYPOFQTCSWLVRV 2106  
 DB 2662 YSQVWVHFVTEHVEHIGFHAKEYFTDCGGIIGDSGVITSPNYPMAYDSLTHCSWLEA 2721  
 QY 2107 EPPDNISLTVYFPLSEKO---YDFELFDGPGSGQSLKALSGNYAPLIYSSNSVY 2162  
 DB 2722 POGHTITLTSDFDIEPHITTCAMDSVYRNGSPESPITIGQYCN--SNPRTIQSGSNQVLV 2780  
 QY 2163 LRWSSDHAAYNRKGFIRYSAPYCSLPRLPHLGFILGOSTOPGSIHFGCNAGRYLVGHS 2222  
 DB 2781 VTFPSDHSLOGGF-----YATWNTQT-----LGGCG---IFHS 2811  
 QY 2223 MALCTRHPGCTHLMSEAIPLCQALS--GCLPEAPRN--GMVPEKEYTVGKAVYSCSEGYHL 2280  
 DB 2812 DNGTIRSBH---WPQNPENSRCSMTAITHKSXHLAISFPNNFLIPSDGDCQCNSPYKV 2867  
 QY 2281 QAGKEALATAC--LDLGLMSNRVNPQCPVPTPDVSSISVEHGRWLLIFETQYQAOQLML 2339  
 DB 2868 WAGTEEVDAKLLATGC---GNVAVG--FVITPSNTFTAV-----POSDBA- 2907  
 QY 2340 ICDGYYVYTGQVRVRCQANGKMSLG-----BSTPCRIIISGCELPFNGHRIG 2388  
 DB 2908 ---PAQGSASFVRCGSNFTGPGSYIISPNYKQYDNNMCTVY-----LEAN----- 2953  
 QY 2389 TLSVYGAITAI--FSCNSGYTLVGSVRCEMANGLWSGSEVRCLAGH-----CG--T 2435  
 DB 2954 PLSTVLLTFVGFHLHARSATVGS-----CVNDG-----VHIRGVSVWSTPATVCGDEM 3003

QY 2436 PEP1-VNGHNGENSVYRGSVVYOCNAGFRLIGMSVRIICQODHHMSGKTPFCVBITCGHP 2494  
 DB 3004 PAPITTINGPULANFYSBQITDF---GFK---FSYHI-----ISCGGV 3040  
 QY 2495 GNPVNGLTQGNQFNLDNVKFCVCPGY--MAEGARSOCLASGQMSDMLPTCRIINCTDPG 2553  
 DB 3041 FNFSSGI-----ITSPAYSADVDPMDHCLVITITVD----- 3072  
 QY 2554 HOENSVQVNASGPHRSFSGTVYSRCNHFYLLGTIVLSCQDGTGTDPRRPOCLLVSCG 2613  
 DB 3073 ---DKVIELKFS-----DFDVVSTSCSHDYLI-----YDGANTSD---PLGKFCG 3114  
 QY 2614 HPGSPHPSQMSG-----DSYTVGAVVRSYCSIGKRTLVGNSTRMCGLDGHWGSLPH 2664  
 DB 3115 SK--RFPVKSNNMSMLVFRTDSDFTQTKMKMF--RQTL--GPOQCG--GYLTG--- 3163  
 QY 2665 CSGTSVGVCDPGI PAHGIRLGDSPDPG-----TYMRPSCENGHVL- 2706  
 DB 3164 ---SNNTPASPDSDSGM-----YDKLNCWIIIAVNVKVIHLTFTNTALBAASTROR 3214  
 QY 2707 -----GSEERTQANGSWSGQPECGVJSCGNPGIPSNARVIFSGLVFSSTIVY 2756  
 DB 3215 CLYDYVVLVDGDSENAMLA--GTFCGSTVPAPFISGNFLT-----VQFISDLTL 3262  
 QY 2757 ECREGVAT--GLASRHC--SVNGTWTG---SDPECLVINGSDPGIPANGRLIGDPRYNK 2810  
 DB 3263 E-REGFNATYITIMDPCGTYNATWIPONISSP-----NSSDPVPS----- 3304  
 QY 2811 TVTYQCVPGYMESHRSVLSCTKDRTNNGTKIPYCKALMCKPPLIPNGKVVGSDPFMWS 2870  
 DB 3305 -----ICTWVIDSPH--QGVKIT---VMAL 3325  
 QY 2871 SVT-YACLEGYOLSLPAVFTCEGNS---WTGELPOCFPVFGCD-----PGYPSR 2916  
 DB 3326 QLTSDQCTQVY---LQLODSDPOGHGNSRFQFCGNASAVVAFYSSMSTAMVIFKSGVNR 3382  
 QY 2917 GREEDRGFSYR--SSVPSCHPPLVLVGSPPRFQSGDSGTGTPSCIDPLTLTCADGVP 2975  
 DB 3383 NSR--MSFTYQIDCNNDYHKAFGNLRSP-----GMPDNYDNDKDCVTTLTA---P 3428  
 QY 2976 Q-----FGIONNSQGYOVGSTVLFRCQKGYL-----LOGSTRTCLPN- 3013  
 DB 3429 QNHTISLFPHSLGENSEVE-----CRNDFLEVRNGSNSNSPLKCYCGTLLPNP 3477  
 QY 3014 -----LWMSGTPDPVPHHCQBPETHYANVAGALDLPBGY 3049  
 DB 3478 VFSQNNELVLRFKSDSVTSDRGYEIIWTSPPSGC-----GGTLGDRGSGFTSP--GY 3527  
 QY 3050 -----TLTPPAR-----ASPS----- 3061  
 DB 3528 PGTYPNNTYCEWVLVAPAGLVTINFYISIDDPGDCVQNYLTLVDGPNASSPSGPGYCG 3587  
 QY 3062 ---RVAPSTAPAR---RMAAGQSRPSAMR 3085  
 DB 3588 GDTSIAPFVASSNQVFIKFIADYARRRSAPR 3618  
 RESULT 3  
 T08618  
 Intrinsic factor-B12 receptor CUBILIN precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 11-Jun-1999 #sequence: revision 11-Jun-1999 #text: change 09-Jul-2004  
 C:Accession: T08618  
 R:Moestrup, S.K.; Koziyaki, R.; Kristiansen, M.; Kayesen, J.H.; Rasmussen, H.H.; Brault, J. Biol. Chem. 273, 5235-5242, 1998  
 A:Title: The intrinsic factor-vitamin B12 receptor and target of teratogenic antibodies  
 A:Reference number: Z16459; MIMD:98148073; PMID:9478979  
 A:Accession: T08618  
 A:Status: preliminary; translated from GB/EMBL/DBUJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3623 <MOE>  
 A:Cross-references: UNIPROT:O70244; EMBL:AF022247; NID:G3834379; PIDN:AACT1661.1; PID:G C:Genetics:

A:Gene: CUBILIN  
 C:Superfamily: Intrinsic factor-B12 receptor cubilin; EGF homology  
 C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membrane  
 F:1-20/Domains: signal sequence #status predicted <SIG>  
 F:21-363/Product: intrinsic factor-B12 receptor CUBILIN #status predicted <MAT>  
 F:133-164/Domains: EGF homology <EGF1>  
 F:436-467/Domains: EGF homology <EGF>

Query Match 8.3%; Score 1404.5; DB 2; Length 3623;  
 Best Local Similarity 20.6%; Pred. No. 9.2e-71;  
 Matches 695; Conservative 378; Mismatches 1083; Indels 1219; Gaps 128;

QY 101 NEGVDLGGSKRTCTKATSDMAFMSDHRPVCRAHMCDAHLRGPGLIT--SPNPPIQYDN 158  
 DB 443 NGGTICIDINGFTC-----DCTSSWTGYCCQTPQAACGILSGTGTAHSPN--DYIH 496  
 QY 159 NAHCWIIITANPKVKIKLAEFPDL-----RGYDTLVGDCGQGDQKTVLYMSQA 212  
 DB 497 NVNCFWIVRT--DEEKVLHVTFTFPDLASASNCPREY--LQIHDDSSADPPLGRY----- 548  
 QY 213 CSDSPHTPGSRIPESMGDIWRQKWTLEICRDISSDARGSVKSPKTSNAVELVAPG 272  
 DB 549 C-----GSRPPGIIH-----SSANALYFLYS 570  
 QY 273 TEIQGSGGDPGIPAYGRREGSRPHHGDYTKFECOPAFELVQKALTCQKNNWSAKPG 332  
 DB 571 EYISGR--GPTA-----RWEALFLE 589  
 QY 333 CVFSCFNFTSPSGVLSPPNVEDYGNHLHCWVILARPESRIHLAINDVER---QF 388  
 DB 590 CGGILTDNY-----GITSPPGYNPPRGDCWQVLVNPNSLITFTGTLISLHNDCK 645  
 QY 389 DFLVIXDGAFAEAFLGTFSGNOLPSSITSGHVARLEFQTDHSTGRGKNITFTT--- 444  
 DB 646 DYLEIRDGPFPQDPVLKFCFTSLSTPLKTTGPAARLHHFSDSTSKGHIITLYLTQSD 705  
 QY 445 -----FHN-----ECDDPGVPVNGKRGFSDLSGSISF 474  
 DB 706 LDCGANTYDDELLPLPSGPFHSRQCVYLITQAQGEQIVINFTHVELSESQMGCSHTY 765  
 QY 475 L-----CDEGLTGGSETITCVLKEGSAVW-----NSAVLRC--EAPCGHLTSPS 519  
 DB 766 IYVODHSLARKICGNETLFPISVSXKWIIRLIDALVOKASFRADYQVACGMARG-B 824  
 QY 520 GTLSPGMPGFYKDALSCAVITAEAPGYPIKITFDREK-----TEVNYDTEVDRGTYS 575  
 DB 825 GFRSPFPYAPYAPGRTCRWTISQPRQVVLNFTPDQIGSSASCDPTDIIEIPSSVLGS 884  
 QY 576 PLGVHGTQVPOFLSTSNVYLIFSTDSHSDIGQLRYETITLQSDHCLDPGLPVNG 635  
 DB 885 PGNBKFSSNIPSFITSVNILLVYTFVKSSMENRGFTAKFSSDKLE----- 931  
 QY 636 QRHGNDFYVVALTFSCDSGYTLSDGEPLCEBNFQSRALPSCALCGGFIGSSGTL 695  
 DB 932 -----CGEVLTAAGTIIIE 944  
 QY 696 SPGPDPFYNNLCTWIETSHKGVFFTFTHLE---SGHDYLLITENGSPFTPLRQ 751  
 DB 945 SPGHFNYPYPRGVNCTMHWVVOQRQLIRLEFFSFFYLEFHVNCTDYLIEYDTAQT-IGR 1003  
 QY 752 LITGSRLPAPISAGLYNGFTAQVAFIDFSMSYEGFNITSEYVLEECBEPVAYIRKG 811  
 DB 1004 YCKKSLPPLSLTS---NSNLIKLFVSDSALAHGFIYNEAID-----ASSV--- 1047  
 QY 812 LQGVGVDTLTFSCEPGYRLEGTRITCLGGRRLMSSPLPRVAECGNSVTGQGTLLSP 871  
 DB 1048 -----CLVYTTDNFGMLSP 1062  
 QY 872 NFPVANNHNECTYISQTOPGKGIQLKAPFELSE-----GDVLKVYDGNNSARLLGV 925  
 DB 1063 NFPNNNPMSWECTYRITTVGLNQIALHFTDFTLEDFYSGQCVDFEIRDDGYETSPLVGI 1122  
 QY 926 FSHSEMAGVTLNSTSSLMWLDFTDAENTSKGFELHSSFELIKCEDPTPRFGKYVME 985

DB 1123 YGCS-VLPPITISHSKMLKFKSDALAKGFSAYW----- 1159  
 QY 986 GHFAGSSVFSQDPGYLSLGSBEILLCLSGERRTWDRPLPTVAECGGTVRGEVSGQVLP 1045  
 DB 1160 ----GSSTG-----CGGNLT--FTGTVLTS 1179  
 QY 1046 GYAPAEHNLNCTWTEAEAGCTIGLFLVPTDEEH-----DVLRIINGPVESSVLLKE 1100  
 DB 1180 NYMPFYHSECEYWRLEASHGSPFELEFQDFILEH--HPSCSLDYLAVFQGPPTNSRLIDK 1238  
 QY 1101 LSGPALPKLHSTFNSVNLQFSTDFTSKQGAIOF-----SVSTATCNDNG 1148  
 DB 1239 LCGDTPAPIRSKVDVLLKATIDAGQQRGEINFRQCDNVVYNTKSGILESINYRN 1298  
 QY 1149 IPONGSRSDSWEA--GDSTVFOCDPGVALQSAEISCVKIE-----NREFWQSPPT 1199  
 DB 1299 PYDKNQRCWMTIQATTGNTVNTYFLGFVDESYNKSTDYVELYDGPQWNGRYCGNNMPRP 1358  
 QY 1200 CIAP-----CGGDLTGPSCVILSPNYPPEYPRKECD 1231  
 DB 1359 GATTGSQLHLVHTPDGINSGEKPKQWFTHCQGGEMSGTAGSPSPGPNYPHNKCI 1418  
 QY 1232 WKVTVSPDYVIALVFNIFNLE-----PGYDFLHIYDGRDSLSPILGSFYG---SOLPGRIE 1284  
 DB 1419 WNIKRVAPGSSIQLTTHDPDVEYHTSCNDLSLEIYAGLDENSPRIHQCSQSPAMPQVS 1478  
 QY 1285 SSSNSLFLAFRSDASVNAAGFVIDYTEMPRESCPFGSIKNGTRVGSDLKLSVTVYCH 1344  
 DB 1479 STGNELAIRFKDSTLNGFGNARAVP----- 1507  
 QY 1345 GGYEVEGTSTLSCIIPDCKPVMNNPRPYCTAPCGGQVYVSDGVLSPPNYPONTSGQIC 1404  
 DB 1508 GG-----CCGIIQLSRGEIHSPPYNNYANMTEC 1536  
 QY 1405 LYFTVPKDYVFGQAFEPH--TALNDVVEVDHGSQHSRLLSLGSHTGESLPLATSNQ 1463  
 DB 1537 SWIIVERRHRLVNLITFDLEAPDSCLRLMDGSSSTNARVASVCGRQOPRSLIASGNS 1596  
 QY 1464 VLKFSAGLAPARBFHVYQAVPRTSATQCSVBEPRYKRLGSDFSVGAIVRECNISG 1523  
 DB 1597 LFRFRSSGSSQNRFPRAEPR----- 1618  
 QY 1524 VALQSPETIECLPVGALAQWNVSAPTCVPCGCGN-LTERGTLISPGFPEPYLNSCV 1582  
 DB 1619 -----ECGGLIMDSSDTIFSPLYPHNYLHNOQS 1648  
 QY 1583 WKI-VVPRGAGIQIOVSP-----VTEQWDSLEVFQDADNTVTMLGSFGTTVALNST 1637  
 DB 1649 WIIEAQPPFNHITLSTFHQLONSSTDCRDEVEILIDNDYDAPVQGRYCGFSLPHRISF 1708  
 QY 1638 SNOLYLFHYSDISVAAGFHELYKTVGLSSCEBPAPVPSNGYKTERLYLVNDVVSITQCEPG 1697  
 DB 1709 GNAITVRVITDSINSEBGFALYSA----- 1733  
 QY 1698 VALQGHAIHSCMPGTVRRRMNYPPLCIAQCGGTVEMEGVILSPGFPGNYSNMDCWKI 1757  
 DB 1734 -----STSSCGGSGYTTLDGINSIDYPRADHPMAECVNI 1768  
 QY 1758 ALPVGFGHIOPLNFSTEP-----NHDYIEINNGEYETSRMKGRPSGSELPSSLST-SHE 1812  
 DB 1769 ASSPENRLQSLFSLNLENSLNCNKDFEVEIREG-NATGHILGRYCNLSLPGNYSABEGHS 1827  
 QY 1813 TTVFPHSDHSONRQGFLEYO-----AY 1835  
 DB 1828 LMRVFSVDSGSGTGNGFQARFKNITPGNNNYVCTHCKIASPFWPKYPNYSNKVWVNDAY 1887  
 QY 1836 -----ELQECDDPEPFANGIVRGAGYN-----VG-----QSYTFECLP 1868  
 DB 1888 HIHGRILMEMDIEPTNCFYSLKIYDGPDRHSRLIGYGCOTQESPSNSNLSLTFQSS 1947  
 QY 1869 GYOLGHPVLTQCGHTRNMDHPKCEVPCGNGITSSNGTV--YSQGFSPYSSSQDCV 1926

```

Db 1948 DSSVSGRFLLEWFAVDSDSTPTIAPGACGGFMWTDGPFVHIFSPGMPREYANGACDI 2007
Qy 1927 WLTTPRGHGVRLNLSTLQTFP-----SGDFTTIDGSGQQAAPRLGVTRSAKTQVSSS 1982
Db 2008 WIIYAP-DSTVELNLIDLEPQOSCNVDKIVXDGSIDSLPVLAVCGVSPPIRSTG 2066
Qy 1983 NOVLKPHRDAATGCIAPALFAVPLTKCP-----PTILPNAVTVENE 2027
Db 2067 EYMIRTSSTSVAG---TGFNASFKHSCGGYIHLADGVTSPKPTDYLPNL-----NC 2118
Qy 2028 EFNIGDIVRYRCLPGLT-----VGNELTCKLQTYLQF-EGP-----PPI----- 2067
Db 2119 SMHV--LVQ-----TGLTIAVHFEQPFQIQRNDSFCGQDVLVLANGDHNSPFLGSGRN 2172
Qy 2068 ---CEVHCPT-----NEL-----TDSTGVI 2085
Db 2173 GRFGGMVAPSTLFTSGNEMFVQFISDSNGQGFIRKREAKSLACGGTVYIHDSDSGYL 2232
Qy 2086 LSQSPGSPYPOFCSTMLVRVEPDYNI SLTVEYFLSEKQYDEF-----ELFP 2132
Db 2233 TSPVYAPNYPOHACITWILEAPPGRSIQLOFE-----DQFNIEDTPNCVSYLEIRD 2284
Qy 2133 GPSQSPBLKALSGNYSAPLIVTSSNSVYLWMSDHAYNRKGFIRYSAPYCSLPAPL 2192
Db 2285 GANSMANLVSILCG-HTLPHSWVSRRRIYLLKFTDGGSSVMGKAYKIASC----- 2236
Qy 2193 HGFLGCTSTOPGSHIFGCNAGRLVGHSMALCTRHPOGYHLMSEALPILCOALSGLPE 2252
Db 2237 ---GGTVSGDSDSVIE---SIGVPTLPYANNV-----FCQMPRIQLP- 2271
Qy 2253 APKRWVFGKEVYTGTAAYVSCSEGYHLOAGAEATACEDLTGTLMSNR-----NV 2301
Db 2372 ---GHYLTLS-----FEDFNLOSFGCTKQFVE--IENHTSGRVLCRYGNS 2414
Qy 2302 PPOCVPTCPDVSSI-----SVHGWRILFETOYQFOQOLMLICDPGYVYTGQRYIRC 2355
Db 2415 TPSSVD-TSSNVAVKFTDGSVYASGRFLQFKSSRQ-----VCG----- 2453
Qy 2356 QANGKMLGD-STYTCRIISCGELPIP---PNQH-RI--GTLSTVGTALF----- 2399
Db 2454 ---GDLHGPT-----GTTSPNYPNPNPHARICEMTIVQGRRLVLTFTNLRL 2499
Qy 2400 ---SCNGYTLVGSRVRECMANGLMSGSEVRCLAGHCGPREPVIHNGENSYRGSV 2455
Db 2500 STOPSNSSEHLIV-----FNQIRNS-----PLQ----- 2524
Qy 2456 VYOCNAGFRLIGMSVRLCOQ-----DHHWSGKTFPCVPITCGHGNPNVGLTQGNQNLN 2510
Db 2525 ---KCSRNVNTEFKSSGNTMKVFFTDG--SRPYGFT----- 2559
Qy 2511 DVYKFCVCPGYMAGAARSCCLASGQWSDMLPCTRIINCTDPGHQENSVR-----QV 2562
Db 2560 ---ASTSTEDAVCGG---FLPSVSGNFPSPGY--NGIRDYANLDCW 2601
Qy 2563 HASGPHRFSTGTVSYRCHNGFYLLGTPLYSCO-----GDGTWDRPRQCLLV 2611
Db 2602 TLSNPNENSISII-----YLELSTISHQCTDVLAFRVGDDGLIEFKCGLSA 2653
Qy 2612 CGHPSPPHSQ-----MSGD--SYTGAVAVYS--CIGKRT--LVNGSTRMCGLDGHW 2658
Db 2654 PTAFLVLPYPCVWHFVNSNERVEYT-GFYIEYFTDGGIRGTGNNGVYISSNPNVLSAW 2712
Qy 2659 TGLSLPHSGSTSVGCGDPGIPAHGIRLGDSPDPTWRFSCGACHVLRGSSERTQANGS 2718
Db 2713 T---HCSWMLKABEG-----HTITL-----TFSDFLLEA-----HPTCTDSV 2747
Qy 2719 W--SGSQPEGVSI--CGNP-----GTPSNARVVFSGLVSSSIYVECEGCVATGLS 2769
Db 2748 TWRIGDPSGSPVIRYCGQSVPRIOGSSNOLIV-----TFNTNOCQCTR-GFYA----- 2796
Qy 2770 RHCSVNGTWTGSDDECLVINCDDPGIIPANGILRLANDFRYNTKVYVQCVGYMESHRSV 2829
Db 2797 ---TWTTN-----ALGCGGTFHSANG-----TISKPHMPQTFPENSRCSW 2833

```

---

```

Qy 2830 LSCFKD-----RTWNGTKPVCKALMC-----KPPILP 2857
Db 2834 TWTHESKWEISPDNSNFRIPSSDSQCQNSFVKWBERLWINKTLATSCGDVAPSTIVT 2893
Qy 2858 NGKRVGSDPFMGGSSVTYACLEGYQLSLPAVFCE--GNGSWTGEIPLQCFPVFCGDPG-- 2912
Db 2894 SGNI-----FTNFOGEBMAQGFASFSIRCGRTNTSPGDI 2931
Qy 2913 -VPSRGREDRGS-----YRSSVSPSCHPLVLVGSPPRRFCQSDG----- 2952
Db 2932 ISPNFPKQYNNMNACTYLIDADPQSLVILTFVSEFHEIDRSGAITGT-----CDHGLHAIK 2987
Qy 2953 ---TWSGTQ---PSCID-PTLTT--CABPGVPOGFIQNNSGCYGVSTVLFRCQK 2998
Db 2988 RNLSSTPLVTICSEYLRPLTVGDPVLNFPYSDAVYTTDFGFKISTYRAITCGG--TYNSS 3045
Qy 2999 GYLQSGSTTRTCLPN 3013
Db 3046 GILRSPSYSNYPN 3060

```

RESULT 4  
136936  
complement receptor 1 - chimpanzee (fragment)  
C/Species: Pan troglodytes (chimpanzee)  
C/Date: 02-Jul-1996 #sequence, revision 02-Jul-1996 #text, change 09-Jul-2004  
C/Accession: 136936, 136937  
R/Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
J. Immunol. 153, 691-700, 1994  
A/Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the 75  
A/Reference number: 136935; MUID:94292799; PMID:8021505  
A/Accession: 136936  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-2014 <BI1>  
A/Cross-references: UNIPROT:Q29530; GB:I24920; NID:g551564; PID:AAA51438.1; PID:g55772  
A/Accession: 136937  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-397,1751-2014 <BI2>  
A/Cross-references: GB:I24921; NID:g557726; PID:AAA51439.1; PID:g557727  
C/Genetics:  
A/Gene: CRI  
C/Superfamily: Cqb-binding protein alpha chain; complement factor H repeat homology  
C/Keywords: duplication; glycoprotein; tandem repeat  
F/18-74/Domain: complement factor H repeat homology <FH01>  
F/79-136/Domain: complement factor H repeat homology <FH02>  
F/141-307/Domain: complement factor H repeat homology <FH03>  
F/213-268/Domain: complement factor H repeat homology <FH04>  
F/272-328/Domain: complement factor H repeat homology <FH05>  
F/336-391/Domain: complement factor H repeat homology <FH06>  
F/336-462/Domain: complement factor H repeat homology <FH07>  
F/470-526/Domain: complement factor H repeat homology <FH08>  
F/529-586/Domain: complement factor H repeat homology <FH09>  
F/722-778/Domain: complement factor H repeat homology <FH10>  
F/1041-1107/Domain: complement factor H repeat homology <FH11>  
F/1172-1228/Domain: complement factor H repeat homology <FH12>  
F/1233-1291/Domain: complement factor H repeat homology <FH13>  
F/1396-1369/Domain: complement factor H repeat homology <FH14>  
F/1432-1489/Domain: complement factor H repeat homology <FH15>  
F/1625-1681/Domain: complement factor H repeat homology <FH16>  
F/1749-1815/Domain: complement factor H repeat homology <FH17>  
F/1823-1879/Domain: complement factor H repeat homology <FH18>  
F/1884-1940/Domain: complement factor H repeat homology <FH19>

Query Match 8.24; Score 1390; DB 2; Length 2014;  
Best Local Similarity 22.24; Pred. No. 2.7e-70;  
Matches 544; Conservative 266; Mismatches 820; Indels 822; Gaps 111;  
798 CEEBEVPA---YSIKRGLQGVGDTLTFSCFPYRLEGTAITLC-LGGRRLWSSPLPRC 853  
Db CRNPDPVNMVHVIDIQ--GSQIKYSTKGYRLIGSSASACIIISGDIVINDNERPIC 136

```

QY 854 -VAECGNSVTGTGTLISPNFVNNNNHCEIYSIQTOPGKIOLKARAFELSGVILKY 912
Db 137 DRICGPLEPTITNDGFISTN---RENHYGVSVTYRCNPGSG---GRKVEFL-VGEP-SI 188
QY 913 YDGNNSARALLGVFSHSEMGMVLTINSTSSLMDFITDAENTSXGFELHFSFELICED 972
Db 189 YCTSNDDQ--VGIMSGPAPQCIIPN-----KCTP 215
QY 973 PGTRFGYKVDDEBHF--AGSSVSFSCDPCGYSLAGSEELCTSGERTTWDRPLEFTCAEC 1030
Db 216 PNVEN-GILVSDNLSLSLNEVVEFRCPQGFVMKGPFRVCKQALNK--MPELEFSCSRVC 272
QY 1031 GGTVRGEVSGQVLSPGYPAPYEHNLMNCIWTIEBAGCTIGLHFLVPTBEVHDVLRIMDG 1090
Db 273 -----QPPP-----DVLH---280
QY 1091 PVESGVLKELISGALPKDHLSTENSYYLOFTDFTSKOGFALQFSVSTATSCNDPGIP 1150
Db 281 -----A 281
QY 1151 QNGSRGDSWEAGDSTVFQCDPGYALOGSAEISGVKIEKPFMPQSPPTCIAOCGDLTG 1210
Db 282 ERTORDXDNFSFGQEVFYSCEPGYDLRGAASLRCTPGD--WSPATPCEVXSCDDFMG 338
QY 1211 P--SGVILSPVPEPPEPPEKCEKWKVTSPDYVIALVFENIFNLEPGYDFLHIYDGRDLS 1268
Db 339 QLNGRVLFP-----VNLQAKXDFVCD-----362
QY 1269 PLIGSPFGSQLPRIESSNSLFLA---FRSDASYNAGFVIDYNTENPRESCEPDGSI 1323
Db 363 -----EGFQLKG--SSASCYVLAMESLMNSVPRCEQIF-----CPSPBPV 402
QY 1324 KNQTRVSDJLKL---GSSVITYYCH---GGYEBGTSITLSCILGDPGKPVNNRPPVC 1374
Db 403 PNQRHTKPLEVPPFGAVVYTCDPHRDGTTFDLIGESTIRCTSDQGGVWSSAPRC 462
QY 1375 TABCGQGVSGDVGLSPNYPNYTSQOICLYFTVVPKDYVVEGFAFHTALNDVVEH 1434
Db 463 -----GIL-----465
QY 1435 DGHSQ--HSRLSSLSGSHGESLPLATSNOLIKTSAKGLADARGF-----HYVOAV 1486
Db 466 -GHQOAPDHFLEFKLKTQTNASDFPIGTS---LTKXCREYVYGRFSTICTLNDLWSS- 519
QY 1487 PR--TSATQCSSYBERPYGK-RLGSDPSVGAIVRECNSSGYALQSGPEICLVPGLAQ 1543
Db 520 PKOVCKRKSCKTPPDVPNGMVHVTIDIOVSRIYNGCTGHRLLGHSSABCI-LSGNAAH 578
QY 1544 MNVSATTCV--VPCGMLTERGTILSPGPEPYLANSICWKLIVBEGAGIOIOVVSFVT 1602
Db 579 WSTKPIQORIPCGLPPTIANGDPISTNENFHYGS-----VVTYNC 620
QY 1603 EQWWSLEVEDGADNVNMLGSFSGTTPALLNSTNOILYHYSDISVSAAGFHELYT 1662
Db 621 NBSRGKRFKFE-----LVGEPSS---LYCTSN-----DQ 646
QY 1663 VGLSSCPER--AVPS-----NGYKTER--YLNDVVSFOCEPGYALOGHAHISC 1709
Db 647 VGIWSPAPQCIIPNCTPPNVENGILVSDNLSLFLNEVVEFRCPQGFVMKGPFRVCKQ 706
QY 1710 PGTVRWMNPPPLCIAQCGGYEEMEGVILSPGFNPYNSMDCSKIALPVGFGAHIOP 1769
Db 707 --ALNMEBELPSCSRVC-----722
QY 1770 LNFSTEPNHXYEIRNGPYETSMGRFSGSELPSLLSHETTYVPHSHSQNRPGFK 1829
Db 723 -----QPPDVL-----HAERTOR-----736
QY 1830 LEYQAVELQECBDDPEPFANGIVRGAGYVNGSGTTECLPGYOLTGHPVLTQCHGTRNMD 1889
Db 737 -----DKDNF-----SPQGEVFPVSCBEGYDLRGAASLRCT--PQGDMS 772

```

---

```

QY 1890 HPLPKCEV--PCGNSITS--SNGVYVSPGPPSPSSODCWMILTVPIGHGVRNLSTLQTE 1947
Db 773 PAAPTCEVXSCDGFMOQLNGRVLFP-----VNLQAKV-----807
QY 1948 PSGDFTIWDGPQQTAPRLGVFTSRMAKKTQSSNOVLLKTHRDAATGITALNFAVAP 2007
Db 808 ---DFVCD--EGFQLKG-----SSASYCVL-----AGMSLMNSVVP 839
QY 2008 LTK--CPEPTILPNAEVATEN--EEFNIGDIYRYCLP-----GFTLVNBIITLKL-- 2055
Db 840 VCEQIFCPSBPVYIPNQRHTGKPLEVFPFGKAVVYTCDPHRDGTTFDLIGESTIRCTSP 899
QY 2056 ---GTLOFEGEPPICEV--HCPTEMLDSTGVLSQSPQSPSTQFOFCSMLVREDEY 2110
Db 900 QGNVW---SSPAPRGILGHCOA-----PD- 922
QY 2111 NISLVEYFLSEKQYDEFEIPDPGQSPPLKALSGNYSAPLITSSNSVYLRWSSDHA 2170
Db 923 -----HFLFAKLKTQTNASDFPIGTSLKYCREYVYGRFSTICTLNDV--WSSPKD 972
QY 2171 YNRKGFKIRYAPYCSLPPAPLHGFILQSTQTOPGSIHFGNAGYRIVGSHMAICTRHP 2230
Db 973 VCR-----KSKCTPPDPVNGMVHVTIDIOVSRIYNGCTGHRLLGHSSABCIISG 1024
QY 2231 QGYHLMSEAIPLCOALSGGLPEAPKGNVFPG--KEYVGTGAVYSCBEGH---LQAG 2283
Db 1025 NSAH--WSTKPIQORIPCGLPPTIANGDPISTNENFHYGVSVTYRCNLSGRKRVFELV 1083
QY 2284 ABATAECL--DTGLMS-----NRNVP-----OCVP 2307
Db 1084 GBSIYCTSNDDQVIMSGPAPQCIIPKCKTPNVENGILVSDNLSLNEVVEFRQF 1143
QY 2308 -----VTC-----PVYSIS-----VEHRMRLIFETQYFOALMLIDP 2343
Db 1144 GFVMKGPFRVCKQALNKMEFELPSCSRVCOPPELHBEHTPSHODNFSPGQEVVYSCBP 1203
QY 2344 GYYTGGQVIRICOANGKWSLGDSTPTCLISG---GELPIPPNQRHTGTLSTV--YGAI 2398
Db 1204 GYDLRGAASLHCTPOGDMS--PEAPRCVXSCDDEFLQD--PHGRVLPPLNLQAKAVS 1258
QY 2399 FSCNSGYTLVSGSRVRECANG--LWSGEVRCLAGHGTPEPYNGHINGENYS---YR 2452
Db 1259 FVCEGFPRLKSSVSHCVLGMRSIMNNSVPRCQIFCPNRPALINGHHTGTPTGDIYFG 1318
QY 2453 GSVYVQCN-----AGFELIGMSYRICQDDH---WSGKTPFC-VPTTCGH-----2493
Db 1319 KEISTTCDPHRDGTTFDLIGESTIRCTSDHNGVWSSPAPRCLELPHAGHCKTPROFP 1378
QY 2494 ---PGNPNVGLTQGNQFNLNDVVFVCNPGYMAEGAARSOCLAGQMSDMLPTCRINCT 2550
Db 1379 FASPTIPIINDF---EPFVGTSLNVECRPGYFGKMFSTIS--CLEMLWMSVSDNCRKSSCG 1433
QY 2551 DPGHDSNVROVHAASGPHRSFGTTSYRCNHGFYLLGTPLVSCQGDG---TMDRPRQ 2607
Db 1434 PEPPEFGMVHIND---TQFGSTVNVYSCIEGFRLLIGSPETTLVAGNNVTYMDKAPIC 1489
QY 2608 LVSQGHGSPPHSGSDSYT-----VCAVVRYS-----IGKTLVGN 2648
Db 1490 EIIIC---EPPTISNGDFSSNNRASFNGTIVTYOCHTGPDGEOLFEVIGERSY--C 1543
QY 2649 TRMCGLDGHWGSLPHCSGTSVGVCGDPCIPIAHGIRLGDG---FDPGTVMRFSCGAGHL 2705
Db 1544 TSXODQVGVWSSPPRCISTN--KCTAPREV--ENAIRPQNRSPFSLTEIYRFRQBPQFV 1600
QY 2706 RGSSEPTCOANGSVSGSQPBCGVISGNP-----GTPSNARVVSFGDLVFSSTIYEC 2758
Db 1601 VGSHTVOCTNGRGPRLPHCSRV--COPPELILHGEHTPSH-----QDNFSPQGEVYVYC 1654
QY 2759 REGYATGSLSRHCSVNGVTWGTGSDPECLVINCDD--PGIANGRLANDFRYKNTYTYOC 2816
Db 1555 EPGYDLRGAASLHCTPOGDMSPEAPRCTVXSCDPLQOLPHGRVLPPLNLQAKAVSFCV 1714
QY 2817 VPGYMWESHRYVS--VLSCTDRTWNGTKPVCKALMCKPPLPLNKGKVVGS---DFMGGSS 2871

```

Db 1715 DEGRFLKRSASHCVLACMK-ALMNSSVPVCEQICRPPAIIILNHRHGTFFGDIPIYKE 1773  
2872 VTYAC-----LEGYQLSLPAVTCBGNWSGTELRQC--PFPVCGDPCVPSRG 2917  
1774 ISYACDTHPRGQMTFNLIG-SSSIRCTSDPQNGWSSPAPRCCLSVPAACPHP--PKIQ 1830  
2918 RREBRG-----FSYRVSFSFCHPPLVVGSPRRFCQSDGTWGTGQSPCIDPTLTTCADP 2972  
1831 NGHDIGHVSILYLPMTTISYICDPCGYLVGKGFICTQGIWSQLDHYCKE---VNCSP 1887  
2973 GVPGFGIQQN---SGQYQVGSSTVLFRCQGYLLQSTTRTCLPMLTWGTPP 3021  
1888 -LFMNGISKELEMKVHYGVYTLCKEDGYTLBESPMWSQCADRMD--PP 1936

RESULT 5  
A43526  
Complement C3d/Epstein-Barr virus receptor 2 precursor - mouse  
N:Alternate names: complement receptor type 2  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: A43526; A43538; A43215; A45802; B32215  
R:Fingeroth, J.D.  
J:Immunol. 144, 3458-3467, 1990  
A:Title: Comparative structure and evolution of murine CR2. The homolog of the human C3d  
A:Reference number: A43526; MUID:90229735; PMID:2139457  
A:Accession: A43526  
A:Molecule type: mRNA  
A:Residues: 1-1025 <F1N>  
A:Cross-references: UNIPROT:P19070; GB:M35684; EMBL:J04153; NID:G192687; PID:AAA37448.1  
R:Fingeroth, J.D.; Benedict, M.A.; Levy, D.N.; Strominger, J.L.  
Proc. Natl. Acad. Sci. U.S.A. 86, 242-246, 1989  
A:Title: Identification of murine complement receptor type 2.  
A:Reference number: A43215; MUID:8908890; PMID:2783485  
A:Accession: A43215  
A:Molecule type: mRNA  
A:Residues: 12-305; 'T', 307-519, 'A', 521-1025 <MOU>  
A:Cross-references: GB:M1132; NID:G192692; PID:AAA63295.1; PID:G192693  
R:Fingeroth, J.D.; Benedict, M.A.; Levy, D.N.; Strominger, J.L.  
Proc. Natl. Acad. Sci. U.S.A. 86, 242-246, 1989  
A:Title: Identification of murine complement receptor type 2.  
A:Reference number: A43215; MUID:8908890; PMID:2783485  
A:Accession: A43215  
A:Molecule type: mRNA  
A:Residues: 343-401; 991-1025 <F12>  
A:Cross-references: GB:J04153  
R:Kurtz, C.B.; Paul, M.S.; Aegerter, M.; Weis, J.J.; Weis, J.H.  
J. Immunol. 143, 2058-2067, 1989  
A:Title: Murine complement receptor gene family. Identification and characterization of  
A:Reference number: A45802; MUID:89381350; PMID:2528587  
A:Accession: A45802  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 'E', 100-101, 292-961, 964-1025 <KUR>  
A:Cross-references: GB:M9281; NID:G192685; PID:AAA37447.1; PID:G387131  
A:Note: the authors failed to translate GGA for residue 421 as Gly, and CCA for residue  
A:Note: the authors translated the codon CAC for residue 727 as Asn  
C:Superfamily: Complement C3d/Epstein-Barr virus receptor; complement factor H repeat hc  
C:Keywords: alternative splicing; duplication; glycoprotein; receptor; transmembrane pro  
F:1-11/Domain: signal sequence #status predicted <SIG>  
F:12-1025/Product: complement C3d/epstein-Barr virus receptor 2 (15-repeat form) #status  
F:12-973/Domain: extracellular #status predicted <EXT>  
F:14-73/Domain: complement factor H repeat predicted <FHO1>  
F:82-136/Domain: complement factor H repeat homology <FHO2>  
F:146-202/Domain: complement factor H repeat homology <FHO3>  
F:207-263/Domain: complement factor H repeat homology <FHO4>  
F:268-334/Domain: complement factor H repeat homology <FHO5>  
F:343-358/Domain: complement factor H repeat homology <FHO6>  
F:402-458/Domain: complement factor H repeat homology <FHO7>  
F:463-514/Domain: complement factor H repeat homology <FHO8>  
F:519-585/Domain: complement factor H repeat homology <FHO9>  
F:594-649/Domain: complement factor H repeat homology <FHO>

F:654-704/Domain: complement factor H repeat homology <FHI2>  
F:709-769/Domain: complement factor H repeat homology <FHI3>  
F:778-833/Domain: complement factor H repeat homology <FHI4>  
F:841-897/Domain: complement factor H repeat homology <FHI5>  
F:902-958/Domain: complement factor H repeat homology <FHI6>  
F:968-989/Domain: transmembrane #status predicted <TM>  
F:990-1025/Domain: intracellular #status predicted <INT>

Query Match 5.3%; Score 906.5; DB 1; Length 1025;  
Best local similarity 27.4%; Pred. No. 2.6e-43;  
Matches 255; Conservative 130; Mismatches 414; Indels 133; Gaps 35;

2183 PYGLPAPLHGLFLLGDTSTPGSSIFGCGAGYLVGSHMAICTRHPGQNLHSEAPL 2242  
26 PYSLP-----IYGVLYLNTYTSPTSLRIGKAFICISQNHATDKAPPI 72

2243 COA-----LSCGLPAPKNGWVFGKE--YVGTAKAVSCSSEGYHLQAGAEATLDTGLW 2296  
73 CESVNTKITSQDPIVPGFMNKGSKAPFRHGDSTFTCKANFTMK--GSKTVMQANEMW 130

2297 SNRNV--PQCVPTCCDVSSISVEGRMRLIFETQIQFOALMLI--CDPGYTTGQRV 2352  
131 GPFLPVCESDPFLECPSLPTIHNGHTGHV---DQFVAGLSVTSCEPGYLLTGKXT 186

2353 IRCQANKWSLGDSTPCRIISGELPIRPNHRIGLTV-YGATAFSCNSGYTVYGR 2411  
187 IKCLSSGDMW--GVTPCKAQCSPKPKFPQGVKELSLQVGTTVYFSCNEGVLQSQOP 244

2412 VRECM--ANGWGSSEVRCLAGHCTPEPIVNGHNG-----ENYSYRGSVYVQCN----- 2460  
245 SSQCVIYQKAIWTKKV--KEILCPRPVPRNASHGSENVYGVSTVYTTDPRPEK 303

2461 -AGRLIGSVYRVC---QODHMSGKTPCV---PITGHP--GNPNGLTQGNQF 2507  
304 GVSFTLLGKTKINTTOSOKTIGWSGAPVCLSTSAVLCLQPKIKRQLSLIK--DSY 361

2508 NINDVYFVCPNGMAAGARSOCLASGOWSDMPTC-----RIINCPDPHOENSV 2559  
362 SYNTVAPSCBPGTLLGNRSIRCNAGTMBPPVPCVKGCQAPKLIIN---GQKDSY 417

2560 ROYVASHPRHFPSTTVYRNCNHFYLLGTPTVLSQCGDGTMDRPPOLVSCGHPSP 2619  
418 LL-----NPDPTSIYSCDPEYLLVGBDTHCTPEGMKPTPTTPCQVACEKPVG--P 468

2620 HSQMSGSYTVGAVVYSCIGKRTLVGNSITRMCGLDGHTGSLPHGSGTSVCGDPCIP 2679  
469 HLFKRPONQFIRTPVNSSCBEGFQLSESAYVLCQGTIPMFEIRLCKEIT---CPPPVI 525

2680 AHGIRLDSFDP---GVVMPFSC---EAGHVLGSSERT-----QANGSWSGSQPEC 2726  
526 HNGHTWSSSEDDVYGVVYVMCPGPEBGKFLIGEOYTHCTSDSRGSGSWSPAPLC 585

2727 GV-----ISGNPCTPSNARVVPDGLVF-SSSIYECREGYATGLSRHCSVNGTGS 2781  
586 KLSLPAVQCTDVHVENVKLTNDKAPRYFVYDVSWFKDDDEYLLSSGQIQCKANNTDPE 645

2782 DPECLVINCDP---GIPNGLRLGNDPFRYKNTVTVQCVGYMESHRSVLSCTYDRT 2837  
646 KPLCKEKG-C-EPKRVHLP-----DSDHILKLVKTCQNGVQLTYTEKCONAENG 696

2838 WNGKRPVCKALMCKPRLINGKVV--SDPMGSSVTVACLEGYVL-----SLPAFTC 2890  
697 WPKKIEVCTVILQCPKXINAGHTGMAGFLYGNVSYECDEGFLYLBKSLQCVNDS 756

2891 EGNQSWTGLPQCFP-----VFCGPVGVPS--RGRREBRGGSYSVVSFCHPPLVLVGS 2944  
757 KGHSGWGGPRPQCLQSSPLNCHDPRVKGHGYKXNKTHSAASHNDIYHFCVNGGIMNGSH 816

2945 RRFQSGDTWSGTPSCIDPTLTTCADPVGVPQGIQNSGQ---YGVSTVLFRCQGY 3000  
817 LIRCHTNVTLPGVPTCIRASLQCSPPSTIPNG--NHTGSIARFPFGMSVWYSCYGF 874

3001 LLQSTTRTGLPMLTWGTPRDCVPHNCRPE 3032

Db 875 IMAGEFARLCTHESGTSQPPRCXKNCSPFE 906

RESULT 6

PL0009

complement C3d/Epstein-Barr virus receptor precursor - human

N/Alternate names: complement receptor 2; CR2/CD21

C/Species: Homo sapiens (man)

C/Date: 30-Jun-1992 Sequence revision 07-Jul-1995 #ext\_change 09-Jul-2004

C/Accession: J10028; A39958; A52036; A24319; B24319; D24319; E24319; F24319; P10028

R/Weis, J.U.; Toofocker, L.E.; Smith, J.A.; Weis, J.H.; Fearon, D.T.

R. Exp. Med. 167, 1047-1066, 1988

A>Title: Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus

A/Accession number: J10028; MUID:86171282; PMID:2832506

A/Molecule type: mRNA

A/Residues: 1-1091 <WE1>

A/Cross-references: UNIPROT:P20023

A/Note: nucleotides 1566-1625 are missing from Figure 1; therefore, residues 522-542 have

R/Moore, M.D.; Cooper, N.R.; Tack, B.F.; Nemerow, G.R.

Proc. Natl. Acad. Sci. U.S.A. 84, 9194-9198, 1987

A>Title: Molecular cloning of the cDNA encoding the Epstein-Barr virus/C3d receptor (complement receptor 2)

A/Accession number: A39958; MUID:86097454; PMID:2827171

A/Molecule type: mRNA

A/Residues: 1-456; 'G', 457-644, 'R', 646-669, 'R', 671-816, 'NCSAEVILKAMILERAP', 835-840, 'L', 841-846

A/Cross-references: GB:003565; NID:G181919; PIDN:AA35784.1; PID:G181920

R/Cujtiak, A.; Harley, J.B.; Frank, M.B.; Gruner, B.A.; Frazier, B.; Holers, V.M.

J. Biol. Chem. 264, 2118-2125, 1989

A>Title: Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus receptor

A/Accession number: A32036; MUID:89123277; PMID:2563370

A/Molecule type: mRNA

A/Residues: 1-456; 'G', 457-658, 718-1050, 'I', 1052-1060, 'E', 1062-1091 <FU>

A/Cross-references: GB:004453

R/Weis, J.U.; Fearon, D.T.; Klickstein, L.B.; Wong, W.W.; Richards, S.A.; De Bruyn Kops, Proc. Natl. Acad. Sci. U.S.A. 83, 5633-5643, 1986

A>Title: Identification of a partial cDNA clone for the C3d/Epstein-Barr virus receptor of complement.

A/Reference number: A94114; MUID:86287311; PMID:3016712

A/Accession: A24319

A/Molecule type: protein

A/Residues: 226-330, 'XILQ', 257-267, 333-341, 583-591, 'Q', 593, 'D', 595-596, 728-735 <WE2>

A/Experimental source: B-lymphoblastoid cell lines SB and Raji

C/Genetics: CR2

A/Gene: GDB:CR2

A/Cross-references: GDB:119802; OMIM:120650

A/Map position: 1q32-1q32

A/Supfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat hc

C/Keywords: alternative splicing; duplication; glycoprotein; transmembrane protein

F:1-20/Domains: signal sequence #status predicted <SIG>

F:21-1091/Product: complement receptor 2 (16-repeat form) #status predicted <MAT1>

F:21-658, 718-1091/Product: complement receptor 2 (15-repeat form) #status predicted <MAT2>

F:91-146/Domains: complement factor H repeat homology <FH01>

F:91-146/Domains: complement factor H repeat homology <FH02>

F:154-210/Domains: complement factor H repeat homology <FH03>

F:215-271/Domains: complement factor H repeat homology <FH04>

F:276-347/Domains: complement factor H repeat homology <FH05>

F:351-406/Domains: complement factor H repeat homology <FH06>

F:410-465/Domains: complement factor H repeat homology <FH07>

F:470-521/Domains: complement factor H repeat homology <FH08>

F:526-592/Domains: complement factor H repeat homology <FH09>

F:601-656/Domains: complement factor H repeat homology <FH10>

F:660-716/Domains: complement factor H repeat homology <FH11>

F:720-772/Domains: complement factor H repeat homology <FH12>

F:777-837/Domains: complement factor H repeat homology <FH13>

F:846-901/Domains: complement factor H repeat homology <FH14>

F:909-965/Domains: complement factor H repeat homology <FH15>

F:970-1026/Domains: complement factor H repeat homology <FH16>

F:1034-1091/Domains: transmembrane #status predicted <TM>

F:1057-1091/Domains: intracellular #status predicted <IM>

F:127, 294, 372, 622, 698, 858, 881, 919/Binding site: carbohydrate (Asn) (covalent) #status

```

Query Match      5.3% Score 904; DB 1; Length 1091.
Best Local Similarity 25.2%; Pred. No.3.9e-43;
Matches 261; Conservative 133; Mismatches 438; Indels 264; Gaps 49;

QY   2012 PPPTLLPNAEVTENEERNIGDIVARYRCLPGTLVGNELTLC-----KL-GTYLDFESGPP    2066
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    26 PEPIL--NGRIISYSTPIAVGTIVIRYSOGSGTFRLIGEKSILLCTIKDKVDGTW---DKPAF    80

QY   2067 ICEVAHCFNELLTBSTGVILISOGRGSAPROQTOSMLVRPEPDVNISLTVEYFLSEKQYD    2126
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    81 KCEY-----FNKKSSCP-----EP                                     94

QY   2127 EEFIFDGSGSGLFKLKLSGNYSARPLIVTSSENVSUWRMSBDHANRKGFKRISAPYCS    2186
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    95 ----IYFG-----GKINGSTPY-----108

QY   2187 LPRALDHGFILOGSTORGGSLNHGCAGRYLGHSNALITRNPOGYHMSEA-IPLCOA    2245
       ::| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    109 -----RHGDSVTPACKTNPSMGNKNSWC-----QAANNMGSPTRLPCVS    148

QY   2246 ISCGLP-EARKGNVFGKEYT-----VGTAUVSCSEGYNLDGAEBATACSLPTGM    2296
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    149 V---PLESCPRLPMIHGHNTSENVGSIAIGSLATYSCEBSGYLV--GEKIINCLSGSKY    203

QY   2297 SNRPVPQCVCYT-----PDVSSISEHGMRMLIFETOFQOAMLICDP    2343
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    204 S-AAPPFCSEBARCKSLGRPENGKVKEBPILIAYGVAN-----FFCDB    244

QY   2344 GYYVVGQRVLICQANGKMKSLEDSTPTFRILISCDELPLRP---NGNIIGTLYSV---YGATA    2397
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    245 GYRLGGPSSHCVALGAGVAMTKPVCELEFC---ESPRLNGRIIGSNLANVASGISLV    301

QY   2398 IFSCHS-----GTVLGSRVRECMAN---GLMSGSEVRC---LAGHGPREPIVNH    2443
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    302 TYTDPRDEEGVNPLLIGESTLRCTVDSOKTGTMGPARGPECISTSAVOCPHPQLRGMM    361

QY   2444 ING--ENVSYSAGSYVYCOCNAGFRLLIGSVARCIOODHNHSGKTREPCVRPICGHRGNVNL    2501
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    362 VSQGKDRTYNDITYIFACMFGEFTLUKSKJOJRCMAOGTWERSABVCK-BQARPNTLNQO    420

QY   2502 TQGN--QENLNDVVKFVCNPGYMABGAARSQCLASGOWSDMLPTCRIINCTDPGHQENS    2558
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    421 KEHHMYRFDPGSTEIKYCNPGYVLVGEEISIQ-TSEVMTRPVQCKVAACEATG----    474

QY   2559 VRQYHASGRRFRSGTTVSVRCNHGNYULLGTPLYSCOGSGOTDRRPCCLLVSCGHRG--    2616
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    475 -ROLLTRQHOF-VREDVNSSCGEGFKLSGSYVQECQGTTFPMFMETRLCKETTCERPVI    532

QY   2617 -SPHQMSGDASYTVGAAVVAVYSCIG-----KTIIVGNSTRMCGLD---GHWTGSLPHC    2665
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    533 YNGAHNTGSSLIEDPFYGTIVTYTNCNPERGERVERSLIGESTICTNSIDGCRGMTSGPARLC    592

QY   2666 SGTSVGV-CDDPGLPAHGIRLGSFPD---GIYMRSCEAGHYLNGSSERTTOANGSMWG    2721
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    593 KLSLLAQQSHVHI-ANGXYISKEARFYENDTVLFKCYSGETFLKGS9DIRCKADNTWP    651

QY   2722 SQPECGLVISGNP-----GPSNARVVFSDGLVFESSIIYEEREGYAVALGILLSRHCSYNG    2776
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    652 EIIPYCEK-GCAPRPGAHLHGHTGCVNF---FVSGMTVDTYDDPGLYLGNKSIHCMMSG    707

QY   2777 TWIGSDPE---CLVINCGDPGLIPANGRLANDFRNKTKVTVQCYAPGYMESHRVSLSC    2832
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    708 NWSASARECETEONHROSIOELIPA-GSFV-----ELWNVNSCODGOULTNHAQOMCD    759

QY   2833 TKDRTWMGTGPVCAKLMCKRPPLIRPKGVUG--SDFMWSSVUTAACLEGYOLSRPNFT    2889
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    760 AEMGINMPKPIPLCKVLIHCRRPVLYVNGKHGTGMAAFNLXGNENVSTECDOGFLLGEEKLO    819

QY   2890 C-----EGNSWTGELPQCF--PVP-CGDPGVPS--RGRREDRGFSYSSVFSFCHPLV    2939
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    820 CRSDSKGHWGSGSPQCLRSRPVTRCRNPEVKGHYKINKTKTHSAASHNDIUYVDNCPGFI    879

```

[illegible]

QY 1443 LLSLSGSHGTEBSLPLATNSQVLIKFSAKGLAPARGFFHVQAVPRTSATQSSQSDPEPRY 1502  
 Db 779 RL-----ISSPK- 785  
 QY 1503 GKRLGSDPSVGATVRFECNSGYALQSDPEIECLPVPGLAQMWNVAPRTCV-----PCGN 1558  
 Db 786 -----PPVGATVQVYCDQGFVLIGSAILFCHDRQAGSPKMSDPAFCLLEQFAPC--- 835  
 QY 1559 LTRRGITLSGFPPEPYLNSLNCWKIVPEGAGIQIQVVSFVTEQMWDSLEVPDADNT 1618  
 Db 836 ----- 835  
 QY 1619 VTMLGSFGTTPVPLNLNSTNQVLIHFVSDISVSAAGFHEKYTKVGLSSCPEPAVPSNGV 1678  
 Db 836 -----H-----GLSA-----PENGA 845  
 QY 1679 KTGERYL--VNDVVSFOCEPGVYALQGHAIHISCMGTVRKMYNPPPLCTA 1725  
 Db 846 RSPERKRLHPAGATTHFSCAPGYVLKQGISICVPGHPHWDPPPLCA 894  
 RESULT 8  
 149540  
 procollagen C-endopeptidase (EC 3.4.24.19) precursor, collagen-like splice form - mouse  
 C.Species: Mus musculus (house mouse)  
 C.Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
 C.Accession: I49540  
 R.Fukagawa, M.; Noboru, S.; Hogan, B.L.M.; Jones, C.M.  
 Dev. Biol. 163, 175-183, 1994  
 A.Title: Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1) which is re  
 A.Reference number: I49540; MUID:944229342; PMID:8174772  
 A.Accession: I49540  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: mRNA  
 A.Residues: 1-991 <RES>  
 A.Cross-references: UNIPROT:P98063; GB:I424755; NID:g433606; PIDN:AAA7306.1; PID:g43960  
 C.Genetics:  
 A.Gene: Bmp-1  
 C.Superfamily: procollagen C-endopeptidase; astacin homology; Clr/Clr repeat homology;  
 C.Keywords: hydrolase; metalloproteinase; zinc  
 F.I135-326/Domain: astacin homology <AST>  
 F.I556-592/Domain: EGF homology <EG1>  
 F.I596-705/Domain: Clr/Clr repeat homology <CLR>  
 F.I712-747/Domain: EGF homology <EG2>  
 F.I218,222,228,277/Binding site: zinc (His, His, His, Tyr) #status predicted  
 F.I219/Active site: Glu #status predicted  
 Query Match 4.4%; Score 755; DB 2; Length 991;  
 Best Local Similarity 25.2%; Pred. No. 9.6e-35;  
 Matches 235; Conservative 126; Mismatches 339; Indels 232; Gaps 29;  
 QY 952 ENTSGKGFELHFSFELIKCEDPGTPKFGYKVHDEGHFAGSVS-----FSCDPGSLNGS 1006  
 Db 244 ENIQPGGEYNFLKMEVGEVESLGET---YDPDSIMHYARNTPFSGCIFLDTIVPKEVANGV 300  
 QY 1007 EELCLCSERTMD-----RPLPTVAEEGGTVRGVSGQVLSPGPARYENHLNLTIM 1059  
 Db 301 KPSI---GQRTRLSKGDIQAARKLYKCPA--GGETLQ--DSTGNFSSPEYPNGYSAMHRCVW 355  
 QY 1060 TTEAAGCTTILHFLVPEPTEV---HDVLRIMGPVRSYGVLKELSGPALPKDLHSTFN 1115  
 Db 356 RISTVPGKILNTSTMDLYSRSLCWYDYVEVRDGFPMKWVNRGFCGKLPPIVSTDS 415  
 QY 1116 SVTLQFSTDFPFTSKQGFALQPSVSTPNTSCNDPGLPQNGSRGSDSWEA--GDSTVFOCDPGY 1174  
 Db 416 RLWVEF-----RSSSNWVGKGFAYVE----- 437  
 QY 1175 ALQGSABEISCVKIENRFPMQSPPTCLAPCGGDLTGSPGVLTLSNYPPEPYPGKCDMKV 1234  
 Db 438 -----ALCGADVKKDKGHIGSPRYPDYTPSKYCIWKRI 470  
 QY 1235 TVSPDYVIALVFNIENEP---GYDFLHYVDGRDLSPLIGSFYQSLQPRIESSNSL 1290

```

Db 471 QVSEGFHVGTLFQSFSEIRHDSCAVDYLEVRDGHSESNLIGRYCGVENPDIDISTSRLL 530
Qy 1231 FLAFPSDASVSNAGVVIDYTNPRESCDPGSINKGRVSGDLKGSVVITY---CHGGY 1347
Db 531 WLKFSVDSISNAGAVNPFKEVDE-CGRP-----NNGGCEORCLNTLGSYKSCDPGY 583
Qy 1348 EVEGSTSLCILGPPGKPVMMNPRVCTAPCGGQVSGDGVLLSPNYPONTTSQICLYF 1407
Db 584 E-----LAPDKR-----CEAACGFLTKLNGSITSPGMPKEPPNKNCTIWQ 625
Qy 1408 VTPKDYVVFQGFAPFHTALNDV-----VEYHDGSHQSRLLSLSGSHGESIPLATSN 1462
Db 626 LVAPTYRISLQDFEFTEGNDVCKYDFEVRSGILTADSKLHGKFCSEKREVI-TSQYN 684
Qy 1463 QVLKFSKAGLAPARGF--HFVYQAVPRTSATQSSVPEPRYGRKLSGDSVGAIVAFEC 1520
Db 685 NMRVEFSDNTSVSKKGFPAHF-----SDKDECS---KONGGQODCVNPFGS-YECQC 734
Qy 1521 NSGYALQGSPEIECLPVPGALQMNVSAPTCVPCGGLTERRGTLISPGFPEPYLSLN 1580
Db 735 RSGFVLHDNKH-DCKE-----ACGCHKVSTSGTITSPMMPDKYPSKE 777
Qy 1581 CVMKIIVPEAGIQLQVVSFVTEQN---WDSLEVPFGADNTVTMLGSEGTVPALLNS 1636
Db 778 CTMAISSTPGHRVXLTFTVEMDIESQPECAVDHLEVPGRDAKAPVLGRFGSKKPEVLA 837
Qy 1637 TSNOLYLHFYSDISVSAAGFHLKYTVGLSCPEPAVPSNGVKTGERIYLVNDVVSFCPE 1696
Db 838 TGNMFLRFYSDNSVQKGFQASHST----- 863
Qy 1697 GYALQGHAIHSCMPGTVRRMNVPPPLCIAOCQGVTE--EMEGVILSPGF-PGNYPSNMD 1752
Db 864 -----ECGGQVRAADVKTDLVYSHAQGDNNVPGV 894
Qy 1753 CSMKIALPVPGGAHIOFLNF---STEPNHDYIEINRGPEYETSRMRGFGSLPSLS 1808
Db 895 CEMVIVAEQYGVLELVQTEVEEETDCGYLELFGDYVSTAPRLGRYCGSGPPEVYS 954
Qy 1809 TSHETTVYFHSQNRPGFKLEYQAYEJOE 1839
Db 955 AGDSVLVRFHSDPTISKGFHLRTYTKFOD 985

RESULT 9
B58788
Procollagen C-endopeptidase (EC 3.4.24.19) precursor, tolloid-like splice form - human
N/Alternate names: bone morphogenic protein 1, tolloid-like splice form
C/Species: Homo sapiens (man)
C/Date: 28-Mar-1998 #sequence_revision 09-Apr-1998 #text_change 09-Jul-2004
C/Accession: AJ37278, B58788
R/Mozney, J.M.; Rosen, V.; Celeste, A.J.; Miltsock, L.M.; Whittiers, M.J.; Kriz, R.W.; Hew
Science 242, 1528-1534, 1998
A/Title: Novel regulators of bone formation: molecular clones and activities.
A/Reference number: AJ37278, PMID:89072730, PMID:3201241
A/Accession: AJ37278
A/Molecule type: mRNA
A/Residues: 1702, EKRPALQPPRGRRPHQLKRVQKRRTPQ<MOZ>
R/Takahara, K.; Lyons, G.E.; Greenspan, D.S.
J. Biol. Chem. 269, 32572-32578, 1994
A/Title: Bone morphogenetic protein-1 and a mammalian tolloid homologue (wtld) are encod
A/Reference number: A58788, PMID:95096114, PMID:7798260
A/Accession: B58788
A/Molecule type: mRNA
A/Residues: 703-986 <TAK>
A/Cross-references: GB:135279; NID:9619860; PIDN:AA61710.1; PID:9619861
C/Genetics:
A/Genes: GDB: BMP1, BMP-1
A/Cross-references: GDB:125203; OMIM:112264
A/Map position: 8p21-8p21
C/Function:
A/Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen type

```

```

C/Superfamily: procollagen C-endopeptidase; astracin homology; C1r/C1s repeat homology;
C/Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication; g
F/1-22/Domain: signal sequence #status predicted <Sig>
F/23-986/Product: procollagen C-endopeptidase tolloid-like splice form #status predicted
F/110-921/Domain: astracin homology <AST>
F/322-431/Domain: C1r/C1s repeat homology <C1R1>
F/435-544/Domain: C1r/C1s repeat homology <C1R2>
F/551-587/Domain: RGF homology <EG1>
F/591-700/Domain: C1r/C1s repeat homology <C1R3>
F/707-742/Domain: EGF homology <EG2>
F/747-856/Domain: C1r/C1s repeat homology <C1R4>
F/860-973/Domain: C1r/C1s repeat homology <C1R5>
F/91,142,332,363,599/Binding site: carbohydrate (Asn) (covalent) #status predicted
F/163-319,185-205,322-348,375-397,435-461,488-510,551-553,559-572,574-587,591-617,644-6
F/213,217,223,272/Binding site: zinc (His, His, His, Tyr) #status predicted
F/214/Active site: Glu #status predicted
F/565,720/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 4.4% Score 746; DB 1; Length 986;
Best Local Similarity 25.0% Pred No. 3,1e-34;
Matches 233; Conservative 125; Mismatches 341; Indels 232; Gaps 29;

Qy 952 EYTSKGFELHFSFELIKCEDPTGPKFYKADBGHFASSVS-----FSCDPGYSLNGS 1006
Db 239 ENIQGOEYVFLKMPQEVESLGET---YDFSIHVAHNTSRGIFLDITVPKYEANGV 295
Qy 1007 EBLCLSGRRTW-----RPLPTVAEGCGTVRGEVSGVLSFGTPAPYEHNLCTIW 1059
Db 296 KPEI---GQRTRLSKGDIQAARKLYKCPA-CEETLQ-DSTGNFSSPEYNGNSAMHGW 350
Qy 1060 TIEBAGCTIGLHFLVPTTEEV---HDVLRIWDPVSSGVLLKELSPALPKDHSFN 1115
Db 351 RISVTPGEKILINFSLDLNLSRLCWMYVEVRDPMKAPLARGFCSSKLEPIVSDS 410
Qy 1116 SVVLQFSTDFTSKQGFALIQSVSSTATSCNDPGLPONGSRGDSNEA-GDSTVFQCDPGY 1174
Db 411 RLWVEF-----RSSSNVWGKGFAYVE----- 432
Qy 1175 ALQGSABISCVKIENRFPWQSPPTCIAPCGGDLTGPSGVLLSPYPERPYRGKCDKV 1234
Db 433 -----AICGADVKKDYGHIQSPYPPDYVPSKCIWRI 465
Qy 1235 TVSPDYVALVFNIFNLPP---GYDFLIHYGRDLSPLIGSFYSQLPRIESSNSL 1290
Db 466 QVSEGFHVGTLFQSFSEIRHDSCAVDYLEVRDGHSESNLIGRYCGYKPKPDIDISTSRLL 525
Qy 1291 FLAFPSDASVSNAGVVIDYTNPRESCDPGSINKGRVSGDLKGSVVITY---CHGGY 1347
Db 526 WLKFSVDSISNAGAVNPFKEVDE-CGRP-----NNGGCEORCLNTLGSYKSCDPGY 578
Qy 1348 EVEGSTSLCILGPPGKPVMMNPRVCTAPCGGQVSGDGVLLSPNYPONTTSQICLYF 1407
Db 579 E-----LAPDKR-----CEAACGFLTKLNGSITSPGMPKEPPNKNCTIWQ 620
Qy 1408 VTPKDYVVFQGFAPFHTALNDV-----VEYHDGSHQSRLLSLSGSHGESIPLATSN 1462
Db 621 LVAPTYRISLQDFEFTEGNDVCKYDFEVRSGILTADSKLHGKFCSEKREVI-TSQYN 679
Qy 1463 QVLKFSKAGLAPARGF--HFVYQAVPRTSATQSSVPEPRYGRKLSGDSVGAIVAFEC 1520
Db 680 NMRVEFSDNTSVSKKGFPAHF-----SDKDECS---KONGGQODCVNPFGS-YECQC 729
Qy 1521 NSGYALQGSPEIECLPVPGALQMNVSAPTCVPCGGLTERRGTLISPGFPEPYLSLN 1580
Db 730 RSGFVLHDNKH-DCKE-----ACGCHKVSTSGTITSPMMPDKYPSKE 772
Qy 1581 CVMKIIVPEAGIQLQVVSFVTEQN---WDSLEVPFGADNTVTMLGSEGTVPALLNS 1636
Db 773 CTMAISSTPGHRVXLTFTVEMDIESQPECAVDHLEVPGRDAKAPVLGRFGSKKPEVLA 832
Qy 1637 TSNOLYLHFYSDISVSAAGFHLKYTVGLSCPEPAVPSNGVKTGERIYLVNDVVSFCPE 1696
Db 833 TGNMFLRFYSDNSVQKGFQASHST----- 858

```

1697 GYALQGHAIHSCMGTYRMMYPPPLCAOGCGTYE---EMEGVILSRGF-PGNYPSNMD 1752  
 859 -----ECGGGVADVDKTKDYLSHAQEDNNYPCGVVD 889  
 1753 CSWKALIPVCGAHIQIQLNF---STERNHYIIRNGPYETSHMGRFSGSELPSSLLS 1808  
 890 CEWIVVAEEGCVGLVQTFFVEEETDCGYIMLFGCYDSTALSLRGYCGSGPREVYS 949  
 1809 TSHETTYFHSDHSQNRPGFLLEYQAYELQE 1839  
 950 AGDSVLVWFHSDDTTKKGFLLRYTSTFKQD 980  
 RESULT 10  
 NBH0H  
 Complement factor H precursor, long splice form [validated] - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text change 09-Jul-2004  
 C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298  
 R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.  
 Biochem. J. 249, 593-602, 1988  
 A:Title: The complete amino acid sequence of human complement factor H.  
 A:Reference number: S00254; MUID:88134059; PMID:2963625  
 A:Accession: S00254  
 A:Molecule type: mRNA  
 A:Residues: 1-1231 <RIP>  
 A:Cross-references: UNIPROT:P08603; EMBL:X00716; NID:g31964; PIDN:CAA68704.1; PID:g31965  
 A:Note: 402-Tyr was also found  
 A:Note: parts of this sequence, including the amino and carboxyl ends of the mature prot  
 R:Estaller, C.; Schwaeble, W.; Dietrich, M.; Weiss, E.H.  
 Eur. J. Immunol. 21, 799-802, 1991  
 A:Title: Human complement factor H: two factor H proteins are derived from alternatively  
 A:Reference number: A60238; MUID:91184292; PMID:1826264  
 A:Accession: A60238  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-56;1177-1231 <EST>  
 A:Note: only portions of this 4.3 kilobase mRNA were sequenced  
 R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.  
 Biochem. J. 201-207, 1987  
 A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human comple  
 A:Reference number: A54726; MUID:88025472; PMID:2889480  
 A:Accession: A54726  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 579-1231 <DAY>  
 A:Cross-references: GB:M17517; NID:g180497; PIDN:AAA52016.1; PID:g180498  
 A:Note: parts of this sequence were determined by protein sequencing  
 R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.  
 Biochem. J. 6, 65-72, 1986  
 A:Title: Partial characterization of human complement factor H by protein and cDNA sequ  
 A:Reference number: A61565; MUID:86188123; PMID:2938641  
 A:Accession: A61565  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1050-1057; T', 1059-1102 <R12>  
 R:Sim, R.B.; Disclipio, R.G.  
 Biochem. J. 205, 285-293, 1982  
 A:Title: Purification and structural studies on the complement-system control protein be  
 A:Reference number: A26505; MUID:83048213; PMID:6215918  
 A:Accession: A26505  
 A:Molecule type: protein  
 A:Residues: 19-20; Q', 22-29; V', 31-33; Q', 35 <SIM>  
 R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P.C.  
 Biochemistry 31, 3626-3634, 1992  
 A:Title: Solution structure of the fifth repeat of factor H: A second example of the con  
 A:Reference number: A44551; MUID:92232649; PMID:1533152  
 R:Norman, D.G.; Barlow, P.N.; Barton, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.  
 J. Mol. Biol. 219, 717-725, 1991  
 A:Title: Three-dimensional structure of a complement control protein module in solution.  
 A:Reference number: A49224; MUID:91278097; PMID:1829116

A:Contents: annotation; NMR structure determination, residues 927-985  
 R:Estaller, C.; Koerfien, V.; Schwaeble, W.; Dietrich, M.P.; Weiss, E.H.  
 J. Immunol. 146, 3190-3196, 1991  
 A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a nove  
 A:Reference number: I56100; MUID:91201892; PMID:1826708  
 A:Accession: I72654  
 A:Status: translated from GB/EMBL/DDB  
 A:Molecule type: mRNA  
 A:Residues: 1047-1231 <RES>  
 A:Cross-references: GB:M65294; NID:g183766; PIDN:AAA35948.1; PID:g183767  
 R:Carroll, J.A.; Bates, R.C.; Smith, A.I.; Tetoz, T.; Arellano, A.; Gordon, D.L.; Burns,  
 Biochim. Biophys. Acta 1289, 305-311, 1996  
 A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.  
 A:Reference number: S66298; MUID:96205365; PMID:8620012  
 A:Accession: S66298  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 411-419;574-578,580-582 <CAR>  
 C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. H  
 C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver  
 C:Genetics: <HF1>  
 A:Gene: GDB:HF1; HF  
 A:Cross-references: GDB:120041; OMIM:134370  
 A:Map position: 1q32-1q32  
 C:Genetics: <HF2>  
 A:Gene: GDB:HF2; HF  
 A:Cross-references: GDB:129095  
 A:Map position: 1q32-1q32  
 A:Note: the correspondence between the two loci and the sequences indicated is unclear;  
 C:Function:  
 A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also incre  
 he alternative complement pathway  
 A:Pathway: complement alternate pathway  
 C:Superfamily: complement factor H; complement factor H repeat homology  
 C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-1229/Product: complement factor H #status experimental <MP>  
 F:19-44/Product: complement factor H, short splice form #status experimental <MAT>  
 F:21-80/Domain: complement factor H repeat homology <FH01>  
 F:85-141/Domain: complement factor H repeat homology <FH02>  
 F:146-205/Domain: complement factor H repeat homology <FH03>  
 F:210-262/Domain: complement factor H repeat homology <FH04>  
 F:246-248/Region: cell attachment (R-G-D) motif  
 F:267-320/Domain: complement factor H repeat homology <FH05>  
 F:325-385/Domain: complement factor H repeat homology <FH06>  
 F:389-442/Domain: complement factor H repeat homology <FH07>  
 F:448-505/Domain: complement factor H repeat homology <FH08>  
 F:509-564/Domain: complement factor H repeat homology <FH09>  
 F:569-623/Domain: complement factor H repeat homology <FH10>  
 F:630-684/Domain: complement factor H repeat homology <FH11>  
 F:691-744/Domain: complement factor H repeat homology <FH12>  
 F:753-803/Domain: complement factor H repeat homology <FH13>  
 F:811-864/Domain: complement factor H repeat homology <FH14>  
 F:870-926/Domain: complement factor H repeat homology <FH15>  
 F:931-984/Domain: complement factor H repeat homology <FH16>  
 F:989-1043/Domain: complement factor H repeat homology <FH17>  
 F:1048-1102/Domain: complement factor H repeat homology <FH18>  
 F:1109-1163/Domain: complement factor H repeat homology <FH19>  
 F:1217-1228/Domain: complement factor H repeat homology <FH20>  
 F:127-66-52-80, 85-129, 114-141, 146-192, 178-205, 210-251, 237-262, 267-309, 294-320, 325-374, 35  
 1-803, 811-853, 839-864, 870-915, 901-926, 931-973, 959-984, 989-1032, 1018-1043, 1048-1091, 1077  
 F:217/Binding site: carboxylate (Asn) (covalent) #status absent  
 F:529, 802, 822, 882, 911/Binding site: carboxylate (Asn) (covalent) #status experimental  
 F:718, 1029, 1095/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 4.2%; Score 721; DB 1; Length 1231;  
 Best Local Similarity 23.1%; Pred. No. 1,1e-31;  
 Matches 314; Conservative 192; Mismatches 516; Indels 340; Gaps 81;  
 1838 QECDEPFPAN-GIVRA---GYNQSVTFECLPGYQLTGHPVLTCQGTNNMD--H 1890  
 19 EDCHLPPRNTTLTSTWSMDQYPTGQAIYKRCPRYSIGNVIMCRG---EWALN 75

```

QY 1891 PLPKCEV-PCG-----GNITSSNGTVYSPGPSPYSSSODCWMLITVPIGHGRNLTS 1942
DB 76 PLRKQKRPCCGHPGRTPTFTLTGQNVFEYGAIVYICNE-----XYOL-- 120
QY 1943 LLQTEPSGDFITWMDGPOOTAIRLGVFRSMKKTVOSSNOVLKFKHDAATGAIPIA 2002
DB 121 -----LGEIN-----YREDDTIG-WTMD 137
QY 2003 FSAVPLTKCPPTLLPNAEVTTE---NEEFNIGDIYRCLPFTLVGNELITCKLCTY 2058
DB 138 IPICEVAKCLPVTAPENKIVASSAMEPDEYHFGQAVAFVNGSYKLEGEEMHCSDGFG 197
QY 2059 LQFEGPPEICEVHCPTEMLLTDSTGVILISQSPGSPQFOTCSMLVAVEPDYNLSLAVEY 2118
DB 198 MSKEPKCV-ELSCKSPVINGS-----PISOKIYKEMERPOYKKNMGEY 243
QY 2119 FLSEKQIDFEIIFPGPSQSPLLKALSGNVSAPLITVSSNSVYLWSSDHAVNRKGFJ 2178
DB 244 ---SERG---DAVCTESGMRPLPECSEKSCDNFYIPNGDYSPLRIK-----HRTGDEI 290
QY 2179 RYS-----APYCSLPRAPL-----HGTILQOTSTORGSGIHF 2211
DB 291 TYOCRNNGFYPATRNTAKCTSTGWIIPAPRCTLKPCDYDITKHGGLYHNNRRPFPVAVG 350
QY 2212 CNAGYRLVGHSMALCTRH---POGY---HL-----WSEAIPLCOALSCLPAPAKNG-- 2257
DB 351 KYYSY-----YCDHFETPSGSYMDHIHCTODGMSPAVPCLR--KCFPPL-ENGYN 399
QY 2258 MVEKEKTVGTKAVYSSSEGYHLOAGAEATACLDITGLMNRNVPOCVPTCPDVSSIS 2317
DB 400 QNHKRAKVOGKSIDVACHPGYALPK-AQTVTTCHEMN-WS---PTPRCLIVKTKCSSSID 454
QY 2318 VEHGRMELIETQYQF--QAOLMLICDPGYYYTGQV---IRCOANGKMSLGDSTPCRI 2372
DB 455 IENG---FISEQYTVALKKAKAYOCTLGVTADGETSGSIRKCKG-WS---AQPTC-I 506
QY 2373 ISCEGLPIPNGRHIGLISY--GATAFSCNSGY-----TLVGSRYECMANGLWSSGE 2425
DB 507 KSC-DIVFNMAKTNDFTWFKLNDITLDECHDGYESNTGSTGIV--CGYNG-WSDLPE 562
QY 2426 VRCLAGHCGTPEPIVNGHI-----NGENYSYRGSVVVQCNAGFLLIMSVARICQDDHNM-- 2479
DB 563 I-CYERCELPK-IDVHLVPRDKKQDYKXGEVAKESKRGFTIVGNVQC---YHFGI 616
QY 2480 SGKTPFCVP--ITCGHFGNPNGLTQ---GNQENLDVYKFCVCPGMAEGARSOCLAS 2534
DB 617 SPDLPICKEQVQSCGPPELLNGVKEKTEBYGHSVEVEYVYCNPRFLMKGPVKIQCV-D 675
QY 2535 GQMSDMLPTCRI--INCTDGHQBNVROVHAGGHRFSGTIVSYRCHGFIYLLCTPVL 2592
DB 676 GEMT-TLPPCTIVESTCGDIPLEHGMAL-SSP-YYYGDSVEFNCSSESFTMIGHRSI 731
QY 2593 SC-----QG----- 2596
DB 732 TCHGWGTQIPQCAIDKLKCKSSNLIIEHLKXKKEPDHNSNIRYCRGKEGHTIV 791
QY 2597 --DGTDRPRPOCLLVS---CGHPSRP--HSQMSGSYTVGAVVRYSICIGKRTLVGNST 2649
DB 792 CINGRMD-PEVNSMAQIQICPPRPQIPNSHMTTLTANRDEGKVSVC--QENYLIQSE 849
QY 2650 RMCGLGHWGSLPHCSGTSVGVCGDPIPAHGI-----RLDSDPDPGTVARFSCGAGHV 2704
DB 850 EITCKXGRMO-SIPLC--VEKIPCSOPPOIHEGTINSSRSQESYAHGFKLSYTEGGGR 906
QY 2705 LRGSSERTQANGSWGSGOPECVLSQNGPTSPNARV--FSDGLVFSSSIYECREGEY 2763
DB 907 ISBENETTCVM-GKMS-SPPQCEGLPCKSPREISHVAVHMSDSYQGEVATYKCFEGRG 964
QY 2764 AAGLSRHSVNGTWTGSDPECLVINGDPGIPA--NGRLGND--FRYNTVTVYQCV 2818
DB 965 IDGPALAK-LGEKMS-HPPSCIKTDC--LSLPSFENALPMGEKQDVYVAGQVTVYTAT 1020

```

```

QY 2819 GYMESHRSVLSCTKDRTWNGTKPYCALMKCPPLINGKVVG--SDFMKSSVTVYA 2875
DB 1021 YTMDDG--ASNVTCLNSR-WTG-RPTCRDYSQVNPPTVQNAIVSRQMSKYPGSEHVRVQ 1076
QY 2876 CLEGYOLSLPAVTCGNGSMWGELPQCFPVF--CGDPCVPSRGREDGFSV--RSSV 2930
DB 1077 CRSPYEMFODEVMCL-NGNWT-EPQCKDSGCKCPPEPINDGITSFPLSYVADASSV 1134
QY 2931 SFSCHPPLVVSPPRFQSDGTSCTOPSCIDPITLTGCADGVPQFGIQNNSQGYQV-- 2988
DB 1135 EYOCQMLYOLEGNKRITCR-NGQMS-EPKCLHPCVIS-----REIMENVIAL 1181
QY 2989 -----GSTVLPFCCKGYLLQSTTCTCPNLWMSG 3018
DB 1182 RWTAKOKLYSRIGSEVFEVCKRGYRL--SSRSHTLRTTCWDG 1221

RESULT 11
T22812
Hypothetical protein ZC116.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T22812, T27494
R/Burton, J.
Submitted to the EMBL Data Library, June 1996
A/Accession: T22812
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3871 <W12>
A/Cross-references: UNIPROT:Q20911; EMBL:Z74473; PDB:CAA98952.1; GSPDB:GN00023; CESP:Z
A/Experimental source: clone F56H9
R/Smyle, R.
Submitted to the EMBL Data Library, June 1996
A/Reference number: Z20376
A/Accession: T27494
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3871 <W12>
A/Cross-references: EMBL:Z74046; PDB:CAA98557.1; GSPDB:GN00023; CESP:ZC116.3
A/Experimental source: clone ZC116
C/Genetics:
A/Gene: CESP:ZC116.3
A/Map position: 5
A/Intons: 29/3; 66/2; 244/1; 359/3; 422/1; 549/1; 572/1; 728/1; 820/1; 889/3; 1189/1;
2763/1; 2809/1; 2906/3; 3051/3; 3176/1; 3254/1; 3344/2; 3536/1; 3565/2; 3582/3; 3609/1;
C/Superfamily: intrinsic factor-B12 receptor cubilin; EGF homology

Query Match 4.0%; Score 685.5; DB 2; Length 3871;
Best Local Similarity 18.8%; Pred. No. 5.8e-30;
Matches 489; Conservative 300; Mismatches 834; Indels 985; Gaps 104,

QY 2 AGAPRALLLPCLSLISDCCASNORHSVG--GSELYKQILEKRGVKKLMPKSDNSQKT 59
DB 1806 ANAPHSALIV-VSTDDKSCGGEVRSOGITVAPFRRIDKPEDC--GVALSNSNGNTVTL 1861
QY 60 SVLTQVGVSGHNNCPDGPGLPERGKRLGSDPRLSSVOFTN-----EGYDQSKRLTCM 115
DB 1862 SIL-----DHNLKSPNCTDSITETIRANSSSKLIRKQCDISSIDSTETESGLYFL 1914
QY 116 K-----VSDMFAAWSDRPVCRAKMDAHLRGP-SGITSPNFPLOYDNA 160
DB 1915 RYRPSASDGDNDDDTDIDQODSNKRPPLFARAYEKVSGRPSKRYVSSP-----MINDTE 1970
QY 161 HCWMTIALNSKYIKLAFBEFDLERGYDITLVGDGQDQDQKTVLYMSQACDSHTP 220
DB 1971 MMVMTLIVDEKAGILVLFID-----LYL-----PSP-TS 1999
QY 221 GSRIPESMGDIWQKWTVEICRDISSSDARSGVAKSPKTSNAVELVAPGTEIEGSGC 280
DB 2000 YLRTFESGEND-----DVASSD-----YEVYGVMAPAKKEFTNL 2035
QY 281 GDPEIPAYRGEGSRFHHGDTLKECQAPAFELVQAKITQCKNNQMSAKKPGCVSCFPN 340

```

```

Db 2036 ----IRVYGLKLEK-----DKFSFTWDPV--FLVYRMLTLAKD-----KXAKYDCC-- 2075
Qy 341 FTSSGVVLSPNYED-----YGNHLHCWLLIARPEERLHAFNDIDVBP- 386
Db 2076 ----GGDLPTFYMEDYFTNPLPAGOSFGEYENLHCWIIIRPMFTGIELFDFYLEDV 2130
Qy 387 ----QFDFLVIR-----DGATAEAPVLGTSGNQLPSISSSGVARLEQTD 430
Db 2131 QNCAPDFVSFRLQDDOPEENDVDVLSVSKCALVNSNNTFNFSVNRALHT--HFVTD 2187
Qy 431 HSTGKGF-----NITFTFRH----- 447
Db 2188 RSRHIGFKLKYRLTCSFEHIRGIFREHTLTSPNYDGNHTPSVWKCOYSLIENRK 2247
Qy 448 -----NECPDPGVVNGKRPD-----SLQSSSISFLCDE 478
Db 2248 VSVEIUDLDIENSPSCYNNALIIGNPFSELVETHSOKSKYCKLEQGERLNFTSAR 2307
Qy 479 GFL-----GTQSEETTCVLEKSGSVWNSAVLRCEAPCGGHLTSPSGTILSPGMPRPYK 532
Db 2308 GRLETKYNSGPNSSKRGFMATIKEQTECPGVLHVDD-----NSPSKVLNSPEFPQRI 2361
Qy 533 DALSCAWIEAOPGYPIKITPDRFKTEVN-----YDLEVRDGRGTYSAPLIGVHGT 584
Db 2362 NSVCEVYMAAPNGHRMLMTFDSNFDIDGNQKCDLDDYIEIRDGPVHSEVIGICGN 2421
Qy 585 QVPOFLISTSNVLYLLPSTDKSHSDIGFOLREYETITLQSDHCLDPGIPVNGQRHNDPV 644
Db 2422 QAPSTIFSTFPLMKLHT-----SEYKSRRF----- 2449
Qy 645 GALVTFSCDSGYTISDGEPLCEBNFQKSRALPSCCALCGFI---QSSGTILSPGPD 701
Db 2450 --VATYGI-----ATCGGTVMREVVTHITSPSBPY 2479
Qy 702 FYPNNLNCTWIIETSHGKGVFTFHTFLBSGHDYLLITENGSTFOPRLQUTGSRLPAP 761
Db 2480 PFALPVGQOMVXKSPN-----THMTLAKYDHWLPLFNPCTMEQLMIR----- 2522
Qy 762 SAGLYNFTAOVRFISDFSMSYEGFNITFSEYDLEPCEEP--VPAYISRKQLQGVGDTL 820
Db 2523 ----DGNSTA-----NPLIGPVCVPRHA-----PDVF 2545
Qy 821 TFGCFPGVRLGSTARITCLGRRRLMSPLRC---VA-----ECGNSVTGOTGLSP 871
Db 2546 TRSASNOITVQFTSNSTTRGRQYCSN--KKCGFDVAVTNSGKCGGRITNLGTLPR 2603
Qy 872 NFPVNNNNHNCIYSIQTPGKGIOLK-----ARAFELSEGDVLKV----- 912
Db 2604 GYPRKLLSHVKCIWDFKMKRGFYTRFKITTPGALAHYHAKFDNQVRIAKQHFPRVAVVEG 2663
Qy 913 ---YDGNNSA-----RLIGVFSHS--EMMGVTINS--TSSSLMLDPLTD 950
Db 2664 FPPYDGAQVSTYFCNKIETQITSITDTIRIVYDSDHTEMISAIGNDQPLNSTFAFPID 2723
Qy 951 ----AENTSGFELHNS-----SPELKCEDPGPK-----F 978
Db 2724 YMALPANSBNGCTLIQINKLITTEPHLANKVNGKGVGLDSFCVHSIEKPANFESVHIEIS 2783
Qy 979 GYKUHDSGHFAGSVSFSFCDPGYSILRGSEELLCSGERRTWDRPLPTCVAE----- 1029
Db 2784 DYKADSELOPAGAK--CFBSMGHLIKISDEPV-----PIETTVCDGKMKMKQS 2830
Qy 1030 -----CGGTVRGVSQVILSPGPA--PYEHN 1054
Db 2831 EMIVNPQIDLFISQLFRDQGOQFNLIIFQCKGCVISBPTGDIISPNGSGIKLPLG 2890
Qy 1055 LNCIWTIEAAGCTIGLHF---LVPTREYHVDLRIWDGPVESGVLLKELSGALPKDL 1110
Db 2891 SKKCMVLEAPGQIVKXIVEMQJAYHDECENHLLIGEG-----ROADVNTIHKYC 2942
Qy 1111 HSTNSVVLQFSTPFTSKOGFALQFSVSTATSCNDPGIPONGSRSGDSW-----EAG 1163

```

---

```

Db 2943 HKMDGE-----OEOKLEDRFKI-----IKSHGRYLLIWTNMKYEAG 2981
Qy 1164 DSTVFOCDPGVALOGSAEISCVKIEENFPWQSPPTCIACGGDLTGSPGVILSPN--PE 1222
Db 2982 -----WKIQYEFLNENE-----BCGHTSOMSGTITYPMGDK 3014
Qy 1223 PYPPGKECDWKVTVSPDYIALVFNIFNL-----PGYDFLHI 1260
Db 3015 DYENDELECTWMDIQVPLGTHNLKFKOPDVTAENCAKDQOLLIOEHSTRANSRPGDYFL 3074
Qy 1261 YDGRDSLPLIGSFYSGQLPGRLESSNSLFLAFRSDASVSNAGVIDYENPRESCFDP 1320
Db 3075 FODEKETPLC-----GIEHPKDESESNRVLNFTTDSQTARGFVNW----- 3119
Qy 1321 GS1KNGTRVSDLKLGSSVTVYCHGYVEVEGTSLSCILDPDGKPVNNRPRVCTARCG 1380
Db 3120 -----EACGA 3125
Qy 1381 QYVSGDGVLSPNYPONYTSGQICLYFVTVP--KQYVFGQAFPHLTALNDVEVHDGSHQ 1439
Db 3126 IYRLNHGVTISPYPNKYPNDITCTYLIARPDDNSVLAIRFADF----- 3169
Qy 1440 HSRLLSLSGSHGESLPLATSNQVLIKFSAKGLAPARG--FHFVQAVERTSAQCSSV 1497
Db 3170 -----DLATMR-----TSFGAPCEDDYLIQIDGTDRVLHTPCASE 3206
Qy 1498 PEPR-----YKRLGSPFSGALVREFCNSGYALQSPLECLFVPGALQOMVNSAP 1549
Db 3207 PMPKDALVFKAIGLKRVSDKS--YYMEEDNKGKGS-----KQPOLAYSIN-- 3253
Qy 1550 TCVPVCGGLTERG-----TILSPGPEYILNSLNCVMKIYVPEGAGIOIYVSVFTEQ 1604
Db 3254 ---KGDNTELBEGTFTITVTSAPFPLTYTDLDCVMMITTDNRQMLIRPEEMLEA 3309
Qy 1605 ---NMDSLVFPDADTVT--MLGSFSGT--TVPALINSTN--QLYLHFPYSDISVSAGF 1656
Db 3310 FHDCSADYIEFPBSSDIMAKTLGKFCGTMDKIPQYRIVTSGPVLILHMTDFVNSGCF 3369
Qy 1657 HLEKTVGLSSCEBPAPVNSGVKTGERYLVNDVVSFQCEBGYALQGHAIHSCMGVTRW 1716
Db 3370 KL-----AVISLQKQEG----- 3382
Qy 1717 NYPRPLCIAQCGG--TVBMEGVILSP--GPPGNYPSMDCSMKIALPVGGAHIOFLNFS 1773
Db 3383 -----CGGKLTITTSWOTLTSBPKDGNIPPALLCGWTISGVDQLOIRIDGV 3432
Qy 1774 TE-----PNHDYLE--INRGPYETSRMGR--FSGSELPSLSLSTSHETVYFH 1818
Db 3433 TEOLEYPRGDRPSEPCIDALAIYDQGFEPBLAAGDICTGTRPLKILYTSHRAPITFE 3492
Qy 1819 SDHSQNRPGFLEYAYEL-----OEC----- 1840
Db 3493 TDBDGTGRGPNISISYDSDCGWLKATNEIKTLVYKIGITSDDNKEMNKERSHQRCFMI 3552
Qy 1841 --PDPEFANGIYRGAGVNYQGVTPRECLGYOLTGHPVLTLCQHTRNMNDHPRKREV- 1897
Db 3553 QGPKTEPV---IYVFOQFNI--PSIAGDCSPFV---EIIVSADIMPKNSRSLKREYN 3605
Qy 1898 ---PCGGIN--TSSNGTVVSPGFPSPYSSSODCVWLITVPIGHVNLNLSLQTEPSGD 1951
Db 3606 ILDSQNRITDNTIKSRLTSPNVPYVYSNSTCV----- 3640
Qy 1952 FITIWDGPQOTAPRLGVFTSRMAKTYOASSNOVLLKFRDAATGIFAIFASVPLTKC 2011
Db 3641 -----TWLQSSNMKML-----LVFNDFTLLE- 3661
Qy 2012 PPTILPNAEVTENEBENI-----GDI--VRYAC--LPQFTLVNELLCKGTIVQFE 2062
Db 3662 -----PNN--VNNKCDYDLMLEKSGDSNGTYCGSLTPALMTSGKDLTALLKSDHSLN 3713
Qy 2063 GPPPICVHCPTE-----LLTDS--TGVLISQSYGSGYPOFOTCSWLV- 2105
Db 3714 HAGYDASYTVVVSBRDQIQFADSYLEGEYISSIGYNGVKNKSYTOFTLIRPNSHDCSI 3773

```

QY 2106 VEPDYNISLT---VEYFLSEKQYDEFEIFDGPSSGSPILKA-----LSGNYSAPLVTYS 2156  
 DB 3774 IFTDVISMTOLPBEPCFPTEBYEIEV---KRTQVKARIRSCFPALKKARELLEA 3829  
 QY 2157 SSNSVYLKMS---SDHAYNRKGFIRY 2180  
 DB 3830 DNDNRVYIKFAFKSDAKSENDGRGFKIRW 3857

## RESULT 12

A45900  
 Complement C3b receptor type 2 long form precursor - mouse (fragment)  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
 C/Accession: A45900; 148306  
 R/Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.  
 J. Immunol. 144, 3581-3591, 1990  
 A/Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2 gene  
 A/Reference number: A45900; MUID:90229754; PMID:2139460  
 A/Accession: A45900  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-676 <KUR>  
 A/Cross-references: UNIPROT:Q99254; GB:M36470  
 R/Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers, V.  
 J. Exp. Med. 181, 151-159, 1995  
 A/Title: Mouse complement regulatory protein Crty/p65 uses the specific mechanisms of bc  
 A/Reference number: 148306; MUID:9510691; PMID:7528766  
 A/Accession: 148306  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 21-367 <RES>  
 A/Cross-references: EMBL:U17128; NID:g595980; PIDN:AA78271.1; PID:g595982  
 C/Genetics:  
 A/Gene: Cr2  
 A/Intons: 80/1, 113/2, 142/1, 275/1, 333/1  
 F/22-78/Domain: complement factor H repeat homology <FH01>  
 F/83-140/Domain: complement factor H repeat homology <FH02>  
 F/145-211/Domain: complement factor H repeat homology <FH03>  
 F/217-272/Domain: complement factor H repeat homology <FH04>  
 F/276-331/Domain: complement factor H repeat homology <FH05>  
 F/336-394/Domain: complement factor H repeat homology <FH06>  
 F/399-458/Domain: complement factor H repeat homology <FH07>  
 F/467-523/Domain: complement factor H repeat homology <FH08>  
 F/531-587/Domain: complement factor H repeat homology <FH09>  
 F/592-648/Domain: complement factor H repeat homology <FH10>

Query Match 4.0%; Score 677; DB 2; Length 676;  
 Best Local Similarity 26.1%; Pired. No. 1.5e-30;  
 Matches 209; Conservative 95; Mismatches 300; Indels 198; Gaps 33;  
 QY 2130 IFDGPSSGSPILKALSGNYSAPLVTSSN-SVYLKMS---SDHAYNRKGFIRY 2180  
 DB 14 IFTGVIGQCKLPLKYS--FAKESIVDKSEPAIGTWYKCRPGFKRSFIITCTETSKW 71  
 QY 2181 SADV-----CSLPAPLHGFLLQGTSTQPGSIFHFGNAGYRLVGHSMALCTRHPOGY 2233  
 DB 72 SDAQCFKRRKPCNMPQEPHGSVHINTGIEFGSTIYSCNQGRIIDGSSAFCIVSDMTV 131  
 QY 2234 HLMSEAIPLCOALSCGLPAPKNGWFGK---EYVGTAYVSCSEGHILQ-----AGAE 2285  
 DB 132 -MNDNDMPCESTPCCSPPAISNGDFYSSRSDFEYGVVVTYCTGKREKLPDLVGEK 190  
 QY 2286 A---TAECLDTGLMNRNVPPOCV-PTCPDVSSISVEHGRWRLIFETGYQQAQMLMIC 2341  
 DB 191 SYTCSKDNQVGMN--SPPOCIRVYKCP---MEIENG-----LV 227  
 QY 2342 DGYVYTGGRVIRCOANGKMSJGDSPTFCRIISCEGLPIPNGHRIGTLVYGATAIRFSC 2401  
 DB 228 ESGFKHS-----FPLND-----TVIRKC 245

QY 2402 NSGYTLVGRVRECMANGLSGSEVRCLAGHCTPEPIYNGHIN--GENYSYRGSVVYQC 2459  
 DB 246 KSGFTWKGSRIRAWCQPSKMSPLPTCFMG-CLPQNIHGHYKNDKDEFFSVQKVSYTC 304  
 QY 2460 NAGFRLIGSVRIICQDDHHWSGKTPCVPIITGHHGNPNVNGLTQGNQF-----NNDVVK 2514  
 DB 305 NRGYTLIGNLVCESTSLGWSNVTVPCEVYKSC--DAIP-NHLHGRVFLPMLQGAEVS 361  
 QY 2515 FVCPNPGYMAEGARSGCLASGC---WSDMLPTCRILINCTPGHNSVQVAGSGPHRS 2571  
 DB 362 FVCDLGFOLKGRPSQCIPEGETVIMNNKFPVCEQISC--DPEPVYNAKPYSLP--IV 418  
 QY 2572 FGTVSYRNKHFFYLLGTPVLSC---QGDGWDPRPOC---LLVSGHFGSPHSGM 2623  
 DB 419 FGTVALRYTSPSYRLIGERALFICISENOVHAATWDAPRCESVNTIISCSDPIVPGGFNM 478  
 QY 2624 SGDS--YTVGAVAVRYSCIGKRTLVNSTRMGLDGHM-TGSLPHSGTSVVGCGPGRIPA 2680  
 DB 479 KSKKAPFRIRGDSVTITCANFTMKSKTYWCQANMGPALPVCESDPLEEC--PSLPT 536  
 QY 2681 --HGIRLG---DSFDPGTVMRPSCEAGHVLGRSSERTCOANGSWGSGQPECGVISCGNPG 2735  
 DB 537 IHNHHTGQHVDFVAGLSVTYSCEPGYLITGKTKIKCLSGMDMGVILPTCKEACERHG 596  
 QY 2736 TFSNARVYFSDGLVYSSIVYECREGYATGLSHNCV---NGTWGSDPECLVINGCD 2792  
 DB 597 KRPNOVKEPISLQVGTIVYFSCNEGLOLQDQSPSCVIVBOKAIWT----- 643  
 QY 2793 PGIPANGRLNDPFRYNTKYVQCVPGYMESHRSVSLCTKDRTNKTKPKVCKALMKCP 2852  
 DB 644 -----KKVVCNEILCP 655  
 QY 2853 PPLINGKTVSGDFMMSGVTY 2874  
 DB 656 PPVVRNGSHTS---FSENVPEY 674

## RESULT 13

NBMSH  
 Complement factor H precursor - mouse  
 N/Alternate names: protein beta-1-H  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 09-Jul-2004  
 C/Accession: A26154; 149711; 149728  
 R/Kristensen, T.; Tack, B.F.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986  
 A/Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length.  
 A/Reference number: A26154; MUID:86233353; PMID:2940596  
 A/Accession: A26154  
 A/Molecule type: mRNA  
 A/Residues: 1-1234 <KRI>  
 A/Cross-references: UNIPROT:P06909; GB:M12660; NID:g193724; PIDN:AAA37759.1; PID:g387181  
 R/Natsunome-Sakai, S.; Nomaoka, M.; Nomaoka, M.; Harada, Y.  
 J. Immunol. 144, 358-362, 1990  
 A/Title: Demonstration of an unusual allelic variation of mouse factor H by the complete  
 A/Reference number: 149711; MUID:90111033; PMID:2136885  
 A/Accession: 149711  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-18 <RES>  
 A/Cross-references: GB:M31979; NID:g193726; PIDN:AAA37762.1; PID:g193729  
 R/Munoz-Canoves, P.; Tack, B.F.; Vlk, D.P.  
 Biochemistry 28, 9891-9897, 1989  
 A/Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma 1r  
 A/Reference number: 149728; MUID:90148935; PMID:2533512  
 A/Accession: 149728  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-19 <RES>  
 A/Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926  
 C/Comment: Two codominant alleles of factor H are present in mice.  
 C/Comment: Factor H functions as a cofactor in the inactivation of C3b by serine protease  
 (C3 convertase) in the alternative complement pathway.

C:Genetics: 1  
 A:Map position: 1  
 C:Keywords: complement factor H; complement factor H repeat homology  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-124/Product: complement factor H #status predicted <MPR>  
 F:21-80/Domain: complement factor H repeat homology <FH01>  
 F:85-141/Domain: complement factor H repeat homology <FH02>  
 F:146-205/Domain: complement factor H repeat homology <FH03>  
 F:210-262/Domain: complement factor H repeat homology <FH04>  
 F:246-248/Region: cell attachment (R-G-D) motif  
 F:325-385/Domain: complement factor H repeat homology <FH05>  
 F:325-385/Domain: complement factor H repeat homology <FH06>  
 F:389-442/Domain: complement factor H repeat homology <FH07>  
 F:448-505/Domain: complement factor H repeat homology <FH08>  
 F:509-564/Domain: complement factor H repeat homology <FH09>  
 F:569-622/Domain: complement factor H repeat homology <FH10>  
 F:629-683/Domain: complement factor H repeat homology <FH11>  
 F:690-743/Domain: complement factor H repeat homology <FH12>  
 F:752-802/Domain: complement factor H repeat homology <FH13>  
 F:808-861/Domain: complement factor H repeat homology <FH14>  
 F:867-931/Domain: complement factor H repeat homology <FH15>  
 F:936-989/Domain: complement factor H repeat homology <FH16>  
 F:1048-1048/Domain: complement factor H repeat homology <FH17>  
 F:1053-1107/Domain: complement factor H repeat homology <FH18>  
 F:1114-1168/Domain: complement factor H repeat homology <FH19>  
 F:1172-1233/Domain: complement factor H repeat homology <FH20>  
 F:121-66,52-80,85-129,114-141,146-152,178-205,210-251,257-262,267-309,294-320,325-374,357-408,808-861,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,1082-1086,721-773,801,1030,1061,1225/Binding site: carbohydrate (Asn) #status pre

Query Match 4.0% Score 676.5; DB 1; Length 1234;  
 Best Local Similarity 23.3%; Pred. No. 3.7e-30;  
 Matches 325; Conservative 155; Mismatches 478; Indels 437; Gaps 81;

1838 QECPEPFPFANG-IVRGAG---YVNGSVTFECLPGYOLGHHVLTCTOHGTNNM--DH 1890  
 19 EDCGPFRESEILSSMSSEQLPBEQTATYKCRPGVRLTGTIVKCKGK---KVVASN 75  
 1891 PLPRC-EVPCCGNITSSNGITVSPGSPSSSODCWLITVPIGHVRLNLSLQTEPS 1949  
 76 PSRIKCRKPCG-----HPG-DTPFGS-----FLAVGSPFEFGAKVYVYCCD 116  
 1950 GDFITINDPGQQTARLGVFTFRSAKKTVOSSNOVL--LKHNDATGCFALAFASAP 2007  
 117 G-----YQLGEIDYRECGADGWINDIPLC--E 142  
 2008 LTKCPPTILPNAEVVT---ENEENIGDIVRYRCLPGFTLVGENEILTKLGTYLQFEG 2063  
 143 VVKCLPTELENGRIVGGAETDQDYFGQVVRFECHSGFKIEBHK----- 188  
 2064 PPICVCHPTNELLTSTGVILSOSYPSGYPOFCQSWLV---RVEPDYNISLTVEYF 2119  
 189 ---EIHCHSENGLSMN-----EKPRCEVILCTPPRVENGGINVKKBYK 228  
 2120 LSEKQVBEFELFDGSPSOSPLKALSGNYSAPLVTSSNSVYLR---WSSDAHYNKGF 2176  
 229 ENERYHYKCKRGVYKERGDAVCTGSGWSQPFCEBKRCSPYLLNGIYPPHRIIHSDD 288  
 2177 KIRYSAPY-----CS-----LPPAPL-----HGFIIGOTSTOP----- 2204  
 289 EIREECYGFYRVGTSTVSKCTPTGWIAPVRCTLKPCFEPQFKGRLYEBSLRPNPVS 348  
 2205 -GGSIHGCGNAGYLVGHSMALCTRHPOGTHLWSEALPLCOALSCGL---PEAR----- 2254  
 349 IGNKYSYKCDNGF-----SPSGYS-W-----DYLRCTAQGMEPEVPCVRKC 389  
 2255 -----KKG--WVEKEVTVGTKAVYSCSEGYHLQAGAEATACLDTGIMSRNVRPCVP 2307  
 390 VFAHYVENGDSAYMKVYVQGSGLVYQCYNGSLNDGQD-TWTCIENG-WS---DPPCIR 444  
 2308 VTCPDVSSISVEHG-----RMRLIETQYQFOALMLTICDPGY-YTCQ--RVRICQA 2357

445 IKTCSASDIHIDNGFLSESSSIYALNRETSYR-----CKQGVYNTGSISSITCLQ 496  
 2358 NGKMS-----LGD----- 2365  
 497 NG-WSPPQSCIKSCDMFVFNISITKTRTPFKLNDKLDYGLVGFENEYKHTGSICTY 555  
 2366 -----STPTCRITISCG-----ELPIPP--NGHRIGTLSVYATAFSCNSGYTLVGRVR 2413  
 556 YGMSDITSCYERECVSUTLTKLVSPRKAKYVGLD-----LEFCHGHR-VGDSV 608  
 2414 ECMANGIWSGEVRC--LACHCGTPEPIYVNGHNGE---NYSYRGSVYVQCMGFRLLIG 2468  
 609 QCHYFG-WSGPFPCCKQVASCAPLEIILNGEINGAKVKEYSHGEVVKYCKCKRFLTKGP 667  
 2469 SVRIQQDHHMSGKTPECV--PITCGHPGPNVGLTGN-----QNLADVYKFCVNP 2550  
 668 NKIQ-CVDGWT-TLPICIEBERTCG---DIELEIGSAKCSVPYHNDVSVEFICEEN 721  
 2521 YMAEGARSGCLASGQW-----SDMLPTCRITINCTDGHQENSRYOVHASGPH--RF 2570  
 722 PTMIGHOSVSCI-SGKMTQLPKCVATDQLEKCYLAKSTG-----IEAIKPKLTER 770  
 2571 SFGITVSYRCNHGFTLLGTPLVLSQCGDGTWRDRPQCL-LVSCGHPGSPPHSQ----- 2622  
 771 THNSTMDYKCRDKQEVERSICI---NGKWD-DEPNCTSKTSCPPPQIENQVIEFTVK 825  
 2623 -MSGDSITVAVVRYSCIGRTLVGNSTRNCGLDGHTGSLPHRCSGTSVGVCGDPGIPAH 2661  
 826 YLQGEKSLVLCQDNY-----LTQSEENVCKDGRWQ-SLPRC--LEKIPCSQPTIEH 875  
 2682 G-IRL-----GDSFDPGTVMRFCEAGHVLVRSSERTQANGSMGSGQPECGV 2728  
 876 GSINLPSSSERDSSHEHGTTSYCDGDFRIPERNRITCYA-GKMS-TPPRCVG 933  
 2729 ISCGNP-----GTPSNARVPSDGLVFSSIVYECREGVYATGLLSHRCVNGTWTGSDP 2783  
 934 LPCGPSPRIPLGTIVLSLELSYQHG---EETVHCSGFGIDGPAFICE-GGKMS-DPP 987  
 2784 ECLVINGDGIIRANGRLG---NDPRYNKTVYQCYPGTGMESHRSVULSCTKDRPWN 2840  
 988 KCIKTDDVLPVYKMAIIRKSKSYRTGQVFRCSPPQNMGS--DYTCVNSR-WIG 1044  
 2841 TKPVCALMKCPPLYPNGKV---GSDPFMGSSVYACLEGYQLSLPAVTFEGNSGWT 2897  
 1045 -QPVCKDNCVDDPPHFNATITVTRTKKUYHGRVRYECKKPYLFGQVWKE-NGIWT 1102  
 2898 GELPQCFPVECGDPGVPSRGRREDRGRFSYRSVSFSCHPPLVVGSPRRFCQSDGTWGT 2957  
 1103 -EKPRKC-----RDSTG-----KCGP----- 1117  
 2958 QPSCIDPTLTTCADPGVPOFGIONNSQGYVGSITVLFRCCKGYLQSTTRTCLPULTW 3017  
 1118 -PRIDNDITSLSLP-----VVEPLSSVEYQCKYVLLKKGKTIITC-TNGKMS 1163  
 3018 GTPPDCVPHHCROPE 3032  
 1164 -BPPTCL-HACVAP 1176

RESULT 14  
 A30359  
 P-selectin precursor - human  
 N:Alternate names: CD62 antigen; granule membrane protein 140  
 C:Species: Homo sapiens (man)  
 C:Date: 18-Oct-1989 #sequence\_revision 30-Sep-1991 #text\_change 16-Aug-2002  
 C:Accession: A30359  
 R:Johnston, G.I.; Cook, R.G.; McEver, R.P.  
 Cell 56, 1033-1044, 1989  
 A:Title: Cloning of GMP-140, a granule membrane protein of platelets and endothelium: s  
 A:Reference number: A30359; PMID:89168432; PMID:2466574  
 A:Molecule type: mRNA  
 A:Residues: 1-830 <JOH>

A:Cross-references: GB:M25322  
 A:Note: parts of this sequence, including the amino end of the mature protein, were conf  
 C:Genetics:  
 A:Gene: GDB:SELP; GRMP  
 A:Cross-references: GDB:120018; OMIM:173610  
 A:Map position: 1q22-1q25  
 C:Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology;  
 C:Keywords: cell adhesion; glycoprotein; phosphohistidine; phosphoprotein; surface anti  
 F:1-41/Domains: signal sequence #status predicted <SIG>  
 F:142-830/Product: F-selectin #status experimental <MAT>  
 F:163-194/Domains: EGF homology <EGF>  
 F:200-257/Domains: complement factor H repeat homology <FH01>  
 F:262-319/Domains: complement factor H repeat homology <FH02>  
 F:324-381/Domains: complement factor H repeat homology <FH03>  
 F:386-443/Domains: complement factor H repeat homology <FH04>  
 F:448-505/Domains: complement factor H repeat homology <FH05>  
 F:510-567/Domains: complement factor H repeat homology <FH06>  
 F:572-629/Domains: complement factor H repeat homology <FH07>  
 F:642-699/Domains: complement factor H repeat homology <FH08>  
 F:704-761/Domains: complement factor H repeat homology <FH09>  
 F:772-795/Domains: transmembrane #status predicted <TM>  
 F:796-830/Domains: intracellular #status predicted <CYT>  
 F:54,98,180,212,219,411,460,518,665,716,723,741/Binding site: carbohydrate (asn) (coval  
 Query Match 3.9%; Score 657.5; DB 2; Length 830;  
 Best Local Similarity 24.3%; Pred. No. 2.5e-29;  
 Matches 196; Conservative 92; Mismatches 325; Indels 193; Gaps 27;

2294 GLMNRNVPPQ---CYPVPCPVSSIVHGRRLIFEQYQFOAQLICDPGYTTG 2349  
 143 GKMDHECKKKKHALCTASQODMS-----SKQGECLFTIGVNTCS---CYPGFY--- 190  
 2350 QRVIRCOANGKMSLGDSTPCRLT-SCGELPIPN-----GHRIGLTVYGATAIRSCNS 2403  
 191 -----GPEVYRECELELPHVLMNSHPLKPS-FNSQSFHCTD 232  
 2404 GYTLVSRVRECMANGMSSEVRCLAGHCCTPEPIVNGH-----NGENYKSGSVVYQC 2459  
 233 GYOVNPSKLECLASITWTKPRQCLACOPPLKIPERGNNICLSAKAFQSSCSFSC 292  
 2460 NAGRLTGMGVRICQODHNSGKTRPCVPIITCGHPGPNV-----LTQGNQFNLNDV 2512  
 293 EBFGLVGPVVOCTASGWTAPAPVCKAVQCOHLEPSEGMDCVNPLT---AFAYGSS 349  
 2513 VKTVCNMGYMAEGARSQCLASGQWSDMLPTCRIINCTDGHQENSVRQVHAS-----GP 2567  
 350 CKFECQGVYVRGLDMRLCIDSGHWSAPLPTCAISC-----EPLESPVHGSMDCSPSL 403  
 2568 HRSFCTTVSYRCHNGRYLLGTPLYLSCQDGTWDRPPQCLLVSCGHPSPHSQNSGDS 2627  
 404 RAYQYDNTCSFRCAEGFMRLGADIVRCDNLGOWTAPVCOALQCODLPPVNEARVN-CS 462  
 2628 YTVGAVVRYSICIKRTLVGNSTRMCGLDGHWGSLPHSCGTSVGVGDGPIRAHGRLDD 2687  
 463 HPEGA-FRYQSV-----CS----- 475  
 2688 SEPFGVVMRFSAGAVLKGSSERTCOANGSWSGSQPECGVISCGNPGTPSNARVVFSDG 2747  
 476 -----FTCNEGILLVGAASYLQCLATGNMNSVPECOAIPCTPLSPQNGTMTCVPR 526  
 2748 L---VFSSSLVYECREGYATATLSRHCSVNGTWTSSDPECLVINGCDGIGIRANGIRLGN 2804  
 527 LGGSSKYSKTCQFLCDGYSLSGPERLDCTRSGRWTDSPMCEAIKPELFAPEQSLDSS 586  
 2805 D---FRYNTVTVYQCVPVGMESHVSVLSCTKDRITNGTKPVCKALCKCKPPLIPNGX 2860  
 587 DTRGEFNVSGTCHFCSCNGFKLEG--PNNVECTSRMSATPTCTCKGASLSTP----- 638  
 2861 VVGSDEMGSSVTYACLEGYQLSLPAVFTCEGNSWTGELPOCFPVYFCSDPGVPSRGRRE 2920  
 639 -----GLQCPAL-TTPGGGT-----MYC-----RHH 658  
 2921 DRGFYSRSSVSPSCHPPLVLVGSPPRFQCSQDGTWSTQPSCTDPLTTCADGCVQFGIQ 2980

Db 659 PGTFGNITCYFCNAGFTLLIGDSTLSCRSPSGWTVTPAC---RAVKSEILVNRPIAM 715  
 Qy 2981 NNSQ---GYOVGSTVLPRCKGYLQGSTTRTCLPVLTVSGTPRDPVPHHCQPERPTH- 3036  
 Db 716 NCSNLMGNYSYSGICGCHLEBGLNGSACTACQENSHMSTVPTQAGRLPTIOEALTYF 775  
 Qy 3037 ----ANVGLDIPSMGYTLITPARR 3057  
 Db 776 GGAVASTIGLI---MGTLTALLRK 797

RESULT 15  
 T31069  
 tolloid-BMP-1 like protein 1 - California sea hare  
 N:Alternate names: probable metalloprotease TBL-1  
 C:Species: Aplysia californica (California sea hare)  
 C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #ext\_change 09-Jul-2004  
 R:Idn, Q.R.; Hattar, S.; Endo, S.; MacPhee, K.; Zhang, H.; Cleary, L.J.; Byrne, J.H.; E  
 U. Neurosci. 17, 765-764, 1997  
 A:Title: A developmental gene (tolloid/BMP-1) is regulated in Aplysia Neurons by treatm  
 A:Reference number: Z20965; MUID:98007484; PMID:8987797  
 A:Accession: T31069  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1070 <LUI>  
 A:Cross-references: UNIPROT:P91972; EMBL:U57369; NID:g1899041; PID:g1899042; PID:AA474  
 C:Superfamily: dorsal-ventral patterning protein tolloid; astacin homology; C1r/C1s repc

Query Match 3.8%; Score 640.5; DB 2; Length 1070;  
 Best Local Similarity 24.2%; Pred. No. 3.3e-26;  
 Matches 222; Conservative 135; Mismatches 331; Indels 231; Gaps 33;

269 VAPGTEIR-----QSGGDP---GIPAYGRREGSRFHNGDTLKECP---AFELV 313  
 332 IMPGKESEFNILDEKVDLSLEPYDYSIMYSHDKFSKSHYLDITRRFRQGMIALPRI 391  
 Qy 314 GQKATTCQKNNQMSAKKGCYFSC--FNFNTPSGCVLSNYPEDYG--NHLCVWLILA 369  
 332 GQNT---RLNDGVDRTNKLKCYKPCGTGWESGITSF---DSGIRGKTCOMRIIA 443  
 370 RPESRIHLAFNDIVER-QPDELVIKDGATAEAPLGFSGNOLPSSITSGHVARLEQ 428  
 444 SHGERIQSLTRLRLSDNCDTIVYEVDRHFPVGLSLGFCCKKLPPIMTSGTRLWVEYK 503  
 Qy 429 TDHSTGKKGFNITFTFRHNECPDPGVVNGKRFDSLQSSISFLCDEGFLTGQSET 488  
 504 S-RARREAFQAAF----- 516  
 489 ITCVLKESVVMNSAVLRCAPPCGHLTSPGTLTSPGMPEGFYKDALSCANVIAORGYR 548  
 517 -----BAICGNMPPGPGFLNSPVPDEYSDKVCCEWIVTVREYQ 557  
 549 IKITFERKTEVN---YDTEVRDGRTYSAPLIGVNGTPOVPOFLISTSNVYLTFSTD 604  
 558 VALERATETETFPDDCAVDYIEIRDDPTKOSPLVGYCGTRTPMAISTSHLVKVPDSD 617  
 Qy 605 KSHDIDIGQLRYETITLQSDHCLDPGIPVNGORHNDYF-VGALVTF--SCDSGYTL-SD 660  
 618 ESMQKGFSAZY---LBEVNEC-----EGEDHGCHVCCNLTGSEYCKIKIGYELHSD 667  
 Qy 661 GEPLCEBNFQMSRALPSCEALGGGFIQSSGTLTSGPDPFYNNMNCWIIETSGKG 720  
 668 GK-----KCEKACGGYLDADBSGTLSSPFDLPPDKNCVWHI--SAPKG 710  
 Db 721 VFTFHTFHL-----ESGHVLLITENGSTFOPLR-QLTGSRLPAPISAGYGNFTAQ 772  
 711 HTLVNFTHMDLEWRGECELDLFRVTVNVGNKERRLQGYCGFMAPPSSITS--LSN-ELR 767  
 Qy 773 VRFISDSMSYEGNITFSEYDLEPCEBEVPAVYSIRKQLQFGVGDTLTSCPPGYNLEG 832  
 768 IEFKSDTLQKTGFSMDYVA-DVDECASSNGGKHCICENT---VG-SFHCSCREGFTIAD 822

QY 833 TARITCGRRRLWSSPLPRCVACGNSVTGOTLLSPNFPVNNNNHCECTYSIQTOPG 892  
Db 823 DEKCKEKG-----CHYEVTDTKGVIQSPDYPSFYPARRDCEMHFTTAPG 867  
QY 893 KGICLKARAFELS-----EGDVLKYDGNNSARLLGVFSHSEMMGVTLNSTSSSLWLD 947  
Db 868 HVRRLIFTDQVEBPHRTCRYDHVEAFDQANIQAPOIGKYGSEKAPIL-SENITLTLTF 926  
QY 948 ITDAENTSKGFELHFSFELIKCEDPGTPKFGYKVHDEGHFAGSSVSFSCDPGYSLRGE 1007  
Db 927 LSDTSVQRKGF-----RARN 941  
QY 1008 ELCLSGBRRTWDRPLPTCVACGTVRGEVSGOVLSPGYAPAYEHNLNCIWTIEAEGC 1067  
Db 942 DTVCOS-----PTAT-----SAPKILSHVLYGS--KPYDNRONCSWNIQAPEGQ 985  
QY 1068 TIGLHPLVFDTEE---VHDVLRIMDGFVESGVLLKELSGPALPKDLHSTFNSVVLQFST 1123  
Db 986 HVELRFTAPEIEQQRCLYDVAVYDGPTEMDLVLGKFCGNQVPSPIVSSTRSLVFRS 1045  
QY 1124 DFTSKOGFAIQFSVSTAT 1142  
Db 1046 DDRIKSGGFSATYRIADDT 1064

Search completed: October 18, 2004, 10:25:19  
Job time : 58 secs

**This page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 18, 2004, 15:00:41 ; Search time 2762 Seconds.

(without alignments)  
5748.994 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985  
Sequence: 1 MAGAPPALLPCLSLSDCC.....RSGVGDPSRLPSSHSPK 3104

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3403857 seqs, 2557783690 residues

Total number of hits satisfying chosen parameters: 6807714

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h  
-Q=/cgn2\_1/USPro.spool/US10016248/runat\_18102004\_105501\_11632/app\_query.fasta\_1.3271  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=tmpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62  
-TRANS=human40.cdi -LIST=45 -DOCALL=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10016248@cgn\_1.1.1858@runat\_18102004\_105501\_11632  
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBBL=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA: \*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq: \*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq: \*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq: \*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq: \*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq: \*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq: \*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq: \*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq: \*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq: \*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq: \*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq: \*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq: \*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq: \*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq: \*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq: \*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq: \*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq: \*  
18: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq: \*  
19: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq: \*  
20: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result

Query

| No. | Score   | Match | Length | DB | ID                  | Description          |
|-----|---------|-------|--------|----|---------------------|----------------------|
| 1   | 16985   | 100.0 | 10136  | 16 | US-10-016-248-1     | Sequence 1, Appl1    |
| 2   | 14142.5 | 83.3  | 8010   | 16 | US-10-016-248-3     | Sequence 3, Appl1    |
| 3   | 9077.5  | 53.4  | 6004   | 16 | US-10-467-433-38    | Sequence 38, Appl1   |
| 4   | 7643.5  | 45.0  | 7323   | 15 | US-10-276-934-5     | Sequence 5, Appl1    |
| 5   | 7643.5  | 45.0  | 8034   | 15 | US-10-276-934-6     | Sequence 6, Appl1    |
| 6   | 7610.5  | 44.8  | 6409   | 15 | US-10-276-934-3     | Sequence 3, Appl1    |
| 7   | 7159    | 42.1  | 5598   | 15 | US-10-276-934-1     | Sequence 1, Appl1    |
| 8   | 6991    | 41.2  | 6145   | 15 | US-10-276-934-2     | Sequence 2, Appl1    |
| 9   | 6978    | 41.1  | 5667   | 15 | US-10-276-934-4     | Sequence 4, Appl1    |
| 10  | 4148.5  | 24.4  | 4506   | 16 | US-10-467-042-27    | Sequence 27, Appl1   |
| 11  | 3942    | 23.2  | 2487   | 9  | US-09-799-514-3     | Sequence 3, Appl1    |
| 12  | 2694    | 15.9  | 2609   | 17 | US-10-399-455-31    | Sequence 69, Appl1   |
| 13  | 2639    | 15.5  | 2387   | 17 | US-10-399-037-69    | Sequence 20, Appl1   |
| 14  | 1940.5  | 11.4  | 14615  | 16 | US-10-451-010-20    | Sequence 31, Appl1   |
| 15  | 1852    | 10.9  | 11230  | 9  | US-09-911-842-3     | Sequence 3, Appl1    |
| 16  | 1852    | 10.9  | 11230  | 13 | US-10-150-821-3     | Sequence 3, Appl1    |
| 17  | 1797.5  | 10.6  | 10878  | 9  | US-09-911-842-1     | Sequence 1, Appl1    |
| 18  | 1797.5  | 10.6  | 10878  | 13 | US-10-150-821-1     | Sequence 1, Appl1    |
| 19  | 1786.5  | 10.6  | 11158  | 15 | US-10-028-248-5     | Sequence 5, Appl1    |
| 20  | 1786.5  | 10.6  | 11158  | 16 | US-10-107-782-5     | Sequence 5, Appl1    |
| 21  | 1793.5  | 10.6  | 11996  | 17 | US-10-603-283-1     | Sequence 7, Appl1    |
| 22  | 1792.5  | 10.6  | 11152  | 15 | US-10-028-248-7     | Sequence 7, Appl1    |
| 23  | 1792.5  | 10.6  | 11152  | 16 | US-10-107-782-7     | Sequence 7, Appl1    |
| 24  | 1782    | 10.5  | 10674  | 15 | US-10-295-027-429   | Sequence 429, Appl1  |
| 25  | 1782    | 10.5  | 10674  | 16 | US-10-295-027-1080  | Sequence 1080, Appl1 |
| 26  | 1638    | 9.6   | 3810   | 16 | US-10-108-260A-814  | Sequence 814, Appl1  |
| 27  | 1616    | 9.5   | 3145   | 16 | US-10-114-270-31    | Sequence 521, Appl1  |
| 28  | 1481    | 8.7   | 9279   | 13 | US-10-087-192-521   | Sequence 521, Appl1  |
| 29  | 1466.5  | 8.6   | 11920  | 17 | US-10-741-601-51    | Sequence 51, Appl1   |
| 30  | 1421.5  | 8.4   | 6153   | 9  | US-09-898-570-21    | Sequence 21, Appl1   |
| 31  | 1421.5  | 8.4   | 6153   | 10 | US-09-839-446-21    | Sequence 21, Appl1   |
| 32  | 1411    | 8.3   | 10872  | 16 | US-10-152-318A-1984 | Sequence 1984, Appl1 |
| 33  | 1371    | 8.1   | 6951   | 16 | US-10-641-1265      | Sequence 1265, Appl1 |
| 34  | 1371    | 8.1   | 6951   | 17 | US-10-283-975A-293  | Sequence 293, Appl1  |
| 35  | 1371    | 8.1   | 7313   | 15 | US-10-037-270-259   | Sequence 259, Appl1  |
| 36  | 1371    | 8.1   | 7313   | 15 | US-10-117-722-259   | Sequence 259, Appl1  |
| 37  | 1370    | 8.1   | 3991   | 16 | US-10-311-623-24    | Sequence 24, Appl1   |
| 38  | 1366.5  | 8.0   | 7028   | 16 | US-10-276-774-802   | Sequence 802, Appl1  |
| 39  | 1314.5  | 7.7   | 1388   | 16 | US-10-416-314-64    | Sequence 64, Appl1   |
| 40  | 1270    | 7.5   | 3804   | 9  | US-09-764-893-43    | Sequence 43, Appl1   |
| 41  | 1270    | 7.5   | 3804   | 9  | US-09-764-881-21    | Sequence 21, Appl1   |
| 42  | 1270    | 7.5   | 3804   | 9  | US-09-764-853-214   | Sequence 214, Appl1  |
| 43  | 1270    | 7.5   | 3804   | 9  | US-09-764-898-60    | Sequence 60, Appl1   |
| 44  | 1270    | 7.5   | 3804   | 10 | US-09-764-881-21    | Sequence 21, Appl1   |
| 45  | 1270    | 7.5   | 3804   | 11 | US-09-764-875-275   | Sequence 275, Appl1  |

#### ALIGNMENTS

RESULT 1  
US-10-016-248-1  
Sequence 1, Application US/10016248  
Publication No. US20040033491A1  
GENERAL INFORMATION:  
APPLICANT: Alabrook et al.  
TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-218  
CURRENT APPLICATION NUMBER: US/10/016,248  
CURRENT FILING DATE: 2002-09-20  
PRIOR APPLICATION NUMBER: 60/254,329  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/291,037  
PRIOR FILING DATE: 2001-05-15  
PRIOR APPLICATION NUMBER: 60/255,648  
PRIOR FILING DATE: 2000-12-14  
PRIOR APPLICATION NUMBER: 60/297,173  
PRIOR FILING DATE: 2001-06-08  
PRIOR APPLICATION NUMBER: 60/309,258  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/326,393  
PRIOR FILING DATE: 2001-10-01

```

; PRIOR APPLICATION NUMBER: 60/315,639
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 10136
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-248-1

Alignment Scores:
Pred. No.: 0 Length: 10136
Score: 16985.00 Matches: 3104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0

US-10-016-248-2 (1-3104) x US-10-016-248-1 (1-10136)

QY 1 MetAlaGlyAlaProProAlaLeuLeuLeuProCysSerLeuLeuIleSerAspCysCys 20
DB 1 ATGGCGGGGGCCCCCTCCCGCCCTTGTCTGTCTGTGCAATTGATCTGAGCTGCTGT 60
QY 21 AlSerAnglnAnghIleSerValGlyValGlyProSerGluLeuValIleYslYsglnIle 40
DB 61 GCTAGCAATCAGGCACTCCGTGGGCGTGAACCTCCGAGCTAGTCAGAAAGCAAAATT 120
QY 41 GluLeuIleSerArgGlyValIleValLeuMetProSerIleYsAspAsnSerGlnIleThrSer 60
DB 121 GAGTTGAAAGTCTGAGGTGTGAGTGTGATGCCAGCAAAAGCAACAGCCGAAAGAGCTCT 180
QY 61 ValLeuThrGlnValGlyValSerGlnGlnIleAsnMetCysProAspProGlyIlePro 80
DB 181 GTGTTAATCAAGTTGGTGTGTGTCTCCAGAGCAATATATGTGTCTCAAGCTTGCATATCCC 240
QY 81 GluArgGlyIleValArgLeuGlySerAspPheArgLeuGlySerSerValGlnPheThrCys 100
DB 241 GAAGAAGGGGCAAAAGACTAGAGCTCGGATTTTCAGGTTAGAGATCCAGCCGTCACCTCG 300
QY 101 AsnGlnGlyIleThrAspLeuGlnGlySerIleYsArgIleThrCysMetIleYsValSerAspMet 120
DB 301 AAGAGAGGCTATACCTCGCAAGGGTCCCAAGCGGATCACCTGTATGAAAGTGAACCCGATG 360
QY 121 PheAlaAlaIlePheSerAspHisArgProValCysArgAlaArgMetCysAspAlaHisIleu 140
DB 361 TTTCGGCCCTCGAGGCAACCAAGCCAGTGTCCGAGCCCGCATGTGTGTATGCCCTT 420
QY 141 ArgGlyProSerArgIleIleThrSerProAsnPheProIleGlnIleThrAspAsnAla 160
DB 421 CGAGGCCCCCTCGGCGATCATCACTCCCAATTTCCCATTCAGATGACAAACATATGCA 480
QY 161 HisCysValIleThrIleThrAlaLeuAsnProSerIleValIleYsIleuAlaPheGlu 180
DB 481 CATCTGTGTGTGATCATCAACAGCACTCAACCTCCCAAGGTGATCAAGCTCGCTTTGAG 540
QY 181 GluPheAspLeuGluArgGlyIleThrAspThrLeuThrValGlyAspGlyIleGlnAspGly 200
DB 541 GACTTTGATTTGAGAGGGGCTATGACACCTTGACCGCTGCGTCAATGTGTGTGATGAG 600
QY 201 AsnGlnIleThrValLeuIleThrSerGlnAsnAlaCysSerAspSerProHisThrPro 220
DB 601 GACCAAAAGCAAGTTCTTCAATGTCTCAAAATGCTCGACATGACAGCCCTCACACCCA 660
QY 221 GlySerArgIleProGluSerMetSerGlyAspIleThrArgGlnIleThrPheThrValLeu 240
DB 661 GGTCTTCGATCCCAAGAGCATGTCTGGAGACATCTGAGGAGGAGAAATGAGCTGTAATT 720
QY 241 GluIleCysArgAspIleSerSerSerAspAlaArgSerGlySerValArgIleSerPro 260
DB 721 GAGATCTGTGTGATGACATTTGAGATTGAGATGCAAGGTTCAGTTCAGTGAAGAAAGTCTCA 780
QY 261 IleThrSerAsnAlaValGluLeuValAlaProGlyThrGluIleGlnGlnIleSerCys 280

```

```

DB 781 AAGACTTAAATGCTGTGAACTTTGTTCTTCGGACAGAAATCAGAGGCACTTCC 840
QY 281 GlyAspProGlyIleProAlaIleThrGlyArgArgGlnGlySerArgPheHisArgIleAsp 300
DB 841 GGTGACCTGTGCAATACCTCATATATGCGGAGGAGAAAGCTCCCGTTTACACCAAGTATG 900
QY 301 ThrIleuIlePheGluIleCysGlnProAlaPheGluLeuValGlyGlnIleValIleThrCys 320
DB 901 ACACTCAAGTTTGAAGTCCAGGCCCGCTTTGAGCTGTGGACAGAAAGCAATCATATCC 960
QY 321 GlnIleYsAsnAnglnIleThrSerAlaIleYslYsProGlyCysValPheSerCysPhePheAsn 340
DB 961 CAAAGAAATACCAATGCTGCGCTTAAGAAAGCAGGCTGCTTTCTTCGCTTCTTCAAC 1020
QY 341 PheThrSerProSerGlyValValIleuSerProAsnIleThrProGluAspIleGlnHis 360
DB 1021 TTCACAGCCCGTCTGGGGTGTCTCTGTCTCCCACTACCCAGAGAGACTATGGCAACAC 1080
QY 361 LeuHisCysValIleThrIleLeuAlaArgProGluSerArgIleHisIleuAlaPheAsn 380
DB 1081 CTTCAGTGTGTGCTCATCTTGGCAGGCTGAGAGCCGATCCACCTGCGCTTCAAC 1140
QY 381 AspIleAspValGluProGlnPheAspPheLeuValIleYsAspGlyValAlaThrAlaGlu 400
DB 1141 GACATTTGAGTGAAGCTCATGTTGATTTCTGTGTCAATCAGATGGGGCCACCCCGAG 1200
QY 401 AlaProValLeuGlyIleThrPheSerGlyAsnGlnLeuAspSerSerIleThrSerSerGly 420
DB 1201 GGGCCCGTCTGGGACCTTCTTCAGAAACCAAGCTTCTCCCTCATCAACAAGCAGTGGC 1260
QY 421 HisValAlaArgLeuGluPheGlnIleThrAspHisSerThrGlyValArgIlePheAsnIle 440
DB 1261 CACGTGGCCCGTCTTGAATTCAGACTCACTCAAGCACTCAGAGGAAAGGGGCTTCAACATC 1320
QY 441 ThrPheThrThrPheArgHisAsnGluCysProAspProGlyValProValAsnGlyIleYs 460
DB 1321 ACTTTTACCACTTCGACACACAGAGTGGCCGATCTGGGGTTCCAGTAAATGGCAAA 1380
QY 461 ArgPheGlnIleAspSerLeuGlnLeuGlySerSerIleSerPheLeuCysAspGluIlePhe 480
DB 1381 CGGTTTGGGAGACAGCTCAGCTGGGCACTTCATCTCTCTGTGATGAGAGCTTC 1440
QY 481 LeuGlyIleThrGlnIleSerGluThrIleThrCysValIleuValGlnGlySerValIleThr 500
DB 1441 CTGGGACTCAGGGCTTCAGAGCACTCACTGCTCTGAAAGAGGGAGGCGGTGTCTGG 1500
QY 501 AsnSerAlaValIleuArgCysGluAlaProCysGlyGlnHisIleuThrSerProSerGly 520
DB 1501 AACAGCGCTGTGCTGCGGTGAGAGCTCCGTGTGTGTCACTTGACCTTCCGCCAGCGGC 1560
QY 521 ThrIleLeuSerProGlyIleThrProGlyIleThrIleYsAspAlaLeuSerCysAlaIleThr 540
DB 1561 ACCATCTCTCTCGGGCTGGCTGTCTTCAAGAGTGGCTTGAAGCTGTGGCTGGGGTG 1620
QY 541 IleGluAlaGlnProGlyIleThrProIleYsIleThrPheAspArgPheIleThrGluVal 560
DB 1621 ATTGAGGCCCAAGCAGGCTTACCCATCAAAATCACTTTCGACAGATTCAAAACGAGGTC 1680
QY 561 AsnIleAspThrLeuGluValAlaArgAspIleArgThrIleThrSerAlaProLeuIleGlyVal 580
DB 1681 AACTATGACACCTCGAAAGTACGCGATGGGAGACTTACTCAGGCGCCCTTATGGGGTT 1740
QY 581 TyrHisGlyIleThrGlnValAlaProGlnPheLeuIleSerThrSerAsnIleThrIleuLeu 600
DB 1741 TACACGGGAGCCAGGTTCCCAAGTCTCTATCAAGCAACCAACATCACTTACTCTCTC 1800
QY 601 PheSerThrAspIleYsSerHisSerAspIleGlyPheGlnLeuValGlyIleThrIleThr 620
DB 1801 TTCTCTACCGACAAGATGATCGGACATGGGCTTCCAGCTCCCTTATGAGACTATTAACA 1860
QY 621 LeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyIleArgHisGlyAsn 640

```

Db 1861 CTGCAGTCAGACCACTGCTGCATCCAGAAATCCAGTAATGACAGCGTCATCGGAAT 1920  
Qy AspPheTyrValGlyAlaLeuValThrPheSerCyAspSerGlyTyrThrLeuSerAsp 660  
Db 1921 GACTTCTACGCGGCGCGCTGGTAGCTTCACGCTGTACCTCGGGCTACACATTAAGGAC 1980  
Qy 661 GlyValProLeuGluCyGluProAspPheGlnTrpSerArgAlaLeuProSerCyGlu 680  
Db 1981 GGGAGGCTCTGGAGGTGAGCCCACTTCAGTGGAGCCGGGCGCTGCCAGTGGAA 2040  
Qy 681 AlaLeuCyGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSerProGlyPhePro 700  
Db 2041 GCTCTCTGTGGTGGCTTCATTCAGAGGCTCCAGTGGAGCACTCTTGCGCAGGGTCCCT 2100  
Qy AspPheTyrProAspAsnLeuAsnCyThrTrpIleIleGlnTrpSerHisGlyValGly 720  
Db 2101 GACTTCAACCCCAACCACTGAACTGCACTCGAATATCGAATACTCATGGCAAGGT 2160  
Qy ValPhePheThrPheHisThrPheHisLeuGluSerGlyHisAspTyrLeuLeuIleThr 740  
Db 2161 GTGTTCTTCACCTTCCACACCTTCACCTGGAAAGTGGCCATGACTACTCTCTCATCT 2220  
Qy GluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760  
Db 2221 GAGAACGGCAGCTTCACTCCAGCCCTGAGGAGCTTAAGTATCTGGATCTCGGCTGCC 2280  
Qy 761 IleSerArgLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780  
Db 2281 ATCAGCCCTGGCTCTATGGCACTTCACTGCCAGGCTCGCTTCATCTCATTTCTCC 2340  
Qy MetSerTyrGluGlyPheAsnIleThrPheSerGlyTyrAspLeuGluProCyGluGlu 800  
Db 2341 ATGTCATATGAAGATTCACATCATCTTCTCAGAGTACGACTTGGAGCCCTGGAGAG 2400  
Qy 801 ProGluValProAlaTyrSerIleArgValGlyLeuGlnPheGlyValGlyAspThrLeu 820  
Db 2401 CCCGAGGCTCCAGCTCAGAGATCCGGAAGGCTTGGAGTTGGCTGGGGGACACTTG 2460  
Qy ThrPheSerCyAspPheProGlyTyrArgLeuGlnGlyThrAlaArgIleThrCyAsnGly 840  
Db 2461 ACCCTTCTCTGCTCCCGGGGTACGCTGGAGGGACCCGCCCATCAGCTGCTGGGG 2520  
Qy 841 GlyArgArgLeuTyrPheSerProLeuProArgCyValAlaGluCyGlyAsnSer 860  
Db 2521 GCGCAGGCGGCGCTGTGAGCTCCCTCTGCCAAGGTGTGTGCTGAGTGGGAATTCA 2580  
Qy 861 ValThrGlyThrGlnGlyThrLeuLeuSerProAspPheProValAsnTyrAsnAsn 880  
Db 2581 GTCAAGGCACTCAGGGATCTTGCTGTCCCAACTTCTCTGGAATCAATTAACAAT 2640  
Qy 881 HisGluCysIleTyrSerIleGlnThrGlnProGlyValGlyIleGlnLeuValAlaArg 900  
Db 2641 CATCAATGCACTCTACTCATCCAGACCAGCAGGAAAGGAAATTCAGCTGAAAGCCAG 2700  
Qy 901 AlaPheGluLeuSerGlyAspValLeuValTyrAspGlyAsnAsnAsnSerAla 920  
Db 2701 GCATTTCAGACTCTCGAAGGAGATGCTCTCAAGGTTATGATGGCAACAACAATCCGCC 2760  
Qy 921 ArgLeuLeuGlyValPheSerHisSerGlyMetMetGlyValThrLeuAsnSerThrSer 940  
Db 2761 CGTTTGTGGAGATTTTAAAGCATTCGAGATGATGGGGGGACTTTGAAACAGCAATCC 2820  
Qy 941 SerSerLeuTrpLeuAspPheIleThrAspAlaGluAsnThrSerValGlyPheGlnLeu 960  
Db 2821 AGCGGTGTGGCTTGAATTCATCATCATGATGAGGACCCAGGAAACCCCAAGTTGGCTAC 2880  
Qy 961 HisPheSerSerPheGluLeuIleValGlySerGluAspProGlyThrProLysPheGlyTyr 980  
Db 2881 CACTTTTCAGCTTTGAATCATCAAAATGAGGACCCAGGAAACCCCAAGTTGGCTAC 2940  
Qy 981 LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCyAspProGly 1000  
Db 2941 AAGGTTATGATGAAGGTCATTTTGGACAGGAGCTCCGTTGCTCTTCACTGTGACCTTGA 3000

Qy 1001 TyrSerLeuArgGlySerGluGluLeuLeuCyLeuSerGlyValAlaArgTrpTrpAsp 1020  
Db 3001 TACAGCTCGGGGATGTAGAGAGCTGCTGTGTCTAGTGTGAGAGCGCCGAGCTGGAG 3060  
Qy 1021 ArgProLeuProThrCyValAlaGluCyGlyGlyThrValArgGlyGluValSerGly 1040  
Db 3061 CGGCTCTGGCCACTGTGTGCGCGAGTGTGGAGGAGACGTAGAGAGAGAGTGTGGGG 3120  
Qy 1041 GlnValLeuSerProGlyTyrProAlaProTyrGlnHisAsnLeuAsnCyIleTrpThr 1060  
Db 3121 CAGGTCTGTCAACCCGGATATCAGGCTCCCTAAGAACAAATCTCAACTGCATCTGAGCC 3180  
Qy 1061 IleGluAlaGluAlaGlyCyThrIleGlyLeuHisPheLeuValPheAspThrGluGlu 1080  
Db 3181 ATGAAACAGAGCCGCTGTGCACATTTGGCTTCACTTCTGTGTGTTGACACAGAGAG 3240  
Qy 1081 ValHisAspValLeuArgIleTrpAspGlyProValGluSerGlyValLeuLeuValGlu 1100  
Db 3241 GTTCAAGAGTGTGGCATCTGGATGGGCTGTGGAGAGGGGGTTCGTGAAGGAG 3300  
Qy 1101 LeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValValLeuGln 1120  
Db 3301 CTGAGTGGCCCGGCTGTGCCAAGACCTGCATAGACACTTCAACTCGTCTGTGAG 3360  
Qy 1121 PheSerThrAspPhePheThrSerValGlnGlyPheAlaIleGlnPheSerValSerThr 1140  
Db 3361 TTCAAGCATGACTTCTTACACAGACAGAGGCTTGGCAATTAATTTCAAGTGTCCACA 3420  
Qy 1141 AlaThrSerCyAsnAspProGlyIleProGluAsnGlySerArgSerGlyAspSerTyr 1160  
Db 3421 GCAAGCTCTGCATATACCTCGGATCCGCAAAATGGAGTGGAGTGTGACAGTTGG 3480  
Qy 1161 GluAlaGlyAspSerThrValPheGlnCyAspProGlyTyrAlaLeuGlnGlySerAla 1180  
Db 3481 GAAGCGGCGCATCCACAGGTTCACAGTGTGACCTTGGCTACCCGCTGCAGGGAAGTGA 3540  
Qy 1181 GluIleSerCyValIleValIleGluAsnArgPhePheThrGlnProSerProProThrCys 1200  
Db 3541 GAGATCAGCTGTGTGAAGTCAGAAACAGTTCTTCTGGAGCCAGCCGCCAACATGTC 3600  
Qy 1201 IleAlaProCyGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr 1220  
Db 3601 ATGCTCTCCGCGGGGAGACCTGCAGAGACCATCTGAGTATCTCTCACCACCAATTAC 3660  
Qy 1221 ProGluProTyrProProGlyValGlyCyAspTrpLysValThrValSerProAspTyr 1240  
Db 3661 CCAGAACCTTACCCGCGCAGGAGAGTGTGACTGAAAGTGAACGCTTCAACAGACTAC 3720  
Qy 1241 ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle 1260  
Db 3721 GTCATGCGCCCTGGATTTTAAACATCTTAACTGTGAACCTGTGGCATCTTCTCCATATC 3780  
Qy 1261 TyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerGlnLeuPro 1280  
Db 3781 TACGAGGAGGAGACTCTCTCAGCCCTCTCATAGAAAGCTTCTATGCTGCCAGCTCCA 3840  
Qy 1281 GlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal 1300  
Db 3841 GCGCGATTGAAAGCAGAGCAACACCTTCTCTGCGCTTCCGACCGCATGATCTGTG 3900  
Qy 1301 SerAspAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCyAspPheAspPro 1320  
Db 3901 AGCAATGTGGCTTGTGATTCATTAACAAGAAACCCGGGAGATCATGTTTGAATCT 3960  
Qy 1321 GlySerIleValAsnGlyThrArgValGlySerAspLeuValLeuGlySerSerValThr 1340  
Db 3961 GGTTCATCAAGAACCGGACACCGGTGGGTCCGACTGAAGTGGGCTCTCTCCATCC 4020  
Qy 1341 TyrTyrCyHisGlyGlyTyrGluValGluGlyThrSerThrLeuSerCyIleLeuGly 1360  
Db 4021 TACTACTGCCACGCGGGCTTACGAAGTTGAGGCACTCGACCTGTGAGCTGACCTTGGAG 4080

|    |      |   |      |
|----|------|---|------|
| Qy | 136  | ProAspGlyIysProValITrpAsnAspProIaGProValCysThrAlaProCysGlyIys     | 1380 |
| Db | 4081 | CCGAGTGGGAAGCCCGGTGGAAACAATCCCGGCCACGTCGTGCACAGCCCCCTGTGGGGGA     | 4140 |
| Qy | 1381 | GlnTyrValGlySerAspGlyValValLeuSerProAsnTyrProGlnAsnTyrThrSer      | 1400 |
| Db | 4141 | CAGATGTGGGCTTCGAGCGSAGTGGCTTGTCTGCCCACTACCCCAAACTACACCACT         | 4200 |
| Qy | 1401 | GlyGlnIleCysLeuTyrPheValIThrValProIysAspTyrValValPheGlyGlnPhe     | 1420 |
| Db | 4201 | GGACACATCGCTGTAATTTTGTACTGTGGCCAAAGACATAGTGTGTGGCCAGTTC           | 4260 |
| Qy | 1421 | AlaPhePheIsthriAlaLeuAsnAspValValGlyValHisAspGlyHisSerGlnHis      | 1440 |
| Db | 4261 | GCCTTCTTTCACAGGCCCTCAACACAGCTGTGTGAGGTTCAAGAGGCGACAGCGCAGC        | 4320 |
| Qy | 1441 | SerArgLeuLeuSerSerLeuSerGlySerHisIsthriGlyGlySerLeuProLeuAlaThr   | 1460 |
| Db | 4321 | TCCGGGCTCCTCAGCTCCCTCTCGGGCTCCCATCAGAGAGATCATGCTCTTGGCCACC        | 4380 |
| Qy | 1461 | SerAsnGlyValLeuIleIysPheSerAlaIysGlyLeuAlaProAlaArgGlyPheHis      | 1480 |
| Db | 4381 | TCCAAICCAAGTTCTCATTAAGTTCAAGGCCAAAGSCTTCGACACAGCGAGGCTTCCAC       | 4440 |
| Qy | 1481 | PheValTyrGlnAlaValProArgThrSerAlaIsthriGlnCysSerSerValProGlnPro   | 1500 |
| Db | 4441 | TTGTGTACCAAGCGGTTCTCGAACACAGCGCCACGAGTCAAGTCTGTGTCCGGAACC         | 4500 |
| Qy | 1501 | ArgTyrGlyLysArgLeuGlySerIAspPheSerValGlyAlaIleValArgPheGlyCys     | 1520 |
| Db | 4501 | CGCTAAGCGAAGAGGTGGGCGTAATCTTCGGTGGGGGCGCATGTCCTTCGAAATGC          | 4560 |
| Qy | 1521 | AsnSerGlyTyrAlaLeuGlnGlySerProGlnIleGlnCysLeuProValProGlyAla      | 1540 |
| Db | 4561 | AACTCCGCTATGCCCTCGAGGGGTGGCCAGAGATCAAGTCCCTCTGTGCTCGGGGCC         | 4620 |
| Qy | 1541 | LeuAlaGlnITrpAsnValSerAlaProThrCysValValProCysGlyGlyAsnLeuThr     | 1560 |
| Db | 4621 | TTGGCCCAATGGAAATGTCACAGGCCACAGTGTGGTGGCTGTGGAGCAACTCACA           | 4680 |
| Qy | 1561 | GluArgArgGlyThrIleLeuSerProGlyPheProGlnProCysLeuAsnSerLeuAsn      | 1580 |
| Db | 4681 | GAGCGAGGGGACACATCTGTGCTCCCTGGCTTCCAGAGCCGTACCTCAACGCTTCAC         | 4740 |
| Qy | 1581 | CysValITrpLysIleValValProGlnGlyValaGlyIleGlnIleGlnValValSerPhe    | 1600 |
| Db | 4741 | TGTGTGTGGAAATGTGTGTCCCGCAAGGCCCTGGATCCAGATCCAAATTGTCAAGTTT        | 4800 |
| Qy | 1601 | ValIThrGlnGlnAsnITrpAspSerLeuGlnValPheAspGlyAlaAspAsnITrpValThr   | 1620 |
| Db | 4801 | GTGACAGAGCACAACTGGGACTCGCTGGAAATATTATGTGTGACGATTAACACTGTAC        | 4860 |
| Qy | 1621 | MetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerAsnGln      | 1640 |
| Db | 4861 | ATGCTGGGGAGTTTCTCAGAAACAACGTCGCTCCCTTCGAACACACCTCAACACG           | 4920 |
| Qy | 1641 | LeuTyrLeuHisIAspTyrSerAspIleSerValSerAlaIleValaGlyPheHisLeuGlyTyr | 1660 |
| Db | 4921 | CTTACCTTCAATTTCTACTCAGATATCAGCTATCTGCACACTGGCTTCCACTTGAGTAC       | 4980 |
| Qy | 1661 | LysThrValGlyLeuSerSerCysProGlnProAlaValProSerAsnGlyValIleIysThr   | 1680 |
| Db | 4981 | AAAAAGGCGGGCCGTGACGATTTGTCCGGAACCTGCTGCGCCAGTAACGGGGTGAAGACT      | 5040 |
| Qy | 1681 | GlyGlyAsnGlyTyrLeuValAsnAspValValSerPheGlnCysGlnProGlyTyrAlaLeu   | 1700 |
| Db | 5041 | GCGCGCGCCCTACTGTGGAATGATGTGTGTCTTTCCAGGTACAGCGGAGTATGCTTC         | 5100 |
| Qy | 1701 | GlnGlyHisAlaHisIleSerCysMetProGlyThrValArgArgITrpAsnTyrProPro     | 1720 |
| Db | 5101 | CAGGCGCACGCGCCACATCTCTCGATGCCCCGGAACAGTGGCGGAGTGAACCTACCTCT       | 5160 |
| Qy | 1721 | ProLeuCysValIleAlaGlnCysGlyGlyThrValGlnGlnMetGlnGlyValIleLeuSer   | 1740 |

|    |      |                    |                    |   |                     |      |
|----|------|--------------------|--------------------|---|---------------------|------|
| Db | 5161 | CCACTCTGATATGCA    | CAGTGTGGGGAGAC     | GTGGAGAGATGGAGGGGCTGATCCTCGAAC          | 522                 |      |
| Qy | 1741 | ProGlyPheProGlyI   | AsnTyrProSerAsnMet | AspCysSerTrpValIleAlaLeuPro             | 1766                |      |
| Db | 5221 | CCGGGCTTCCAGGCA    | CACTACCCAGTAA      | CAATGAGACTGCTCTGGAAAAATACACATCGGCC      | 5281                |      |
| Qy | 1761 | ValGlyPheGlyVal    | IleHisIleGlnPheLeu | AsnPheSerThrGlyProLeuHisAspTyr          | 1781                |      |
| Db | 5281 | GTGGGCTTTGGAGCT    | CACATCCAGTTCTCG    | GAATCTTCCACCGAGGCCCAACCA                | CAGACTAC            | 5341 |
| Qy | 1781 | IleGlnIleArgAsnGly | ProTyrGlnThrSer    | ArgMetMetGlyArgPheSerGlySer             | 1801                |      |
| Db | 5341 | ATAGAATCCGGAAATG   | CGCCCTCATAGAC      | ACAGCCCACTGATGGAAATTCAGTGGAGAC          | 5401                |      |
| Qy | 1801 | GluLeuProSerSer    | LeuDeuSerThrSerIle | GluThrThrValTyrPheHisSerAsp             | 1821                |      |
| Db | 5401 | GAGCTTCCAAAGCT     | CTCTCTCTCCACGTC    | CCACGAGCACCCGTGATTTCCACAGCGAC           | 5461                |      |
| Qy | 1821 | HisSerGlnAsnArg    | ProGlyPheLysLeuGln | TyrGlnAlaTyrGluLeuGlnGlyCys             | 1841                |      |
| Db | 5461 | CAGTCCCAAGAAATG    | CGGCAAGGATTCAG     | GCTGAGGTATCAGGCTTAAGAACTTCAAGAGTGC      | 5521                |      |
| Qy | 1841 | ProAspProGlnPro    | PheHlaAsnGlyIle    | ValArgGlyValAspTyrAsnValGlyGln          | 1866                |      |
| Db | 5521 | CCAGACCCAGAGCC     | CTTTGGCCAAATGCA    | TTGTGAGGGAGCGCTGACAACTGGAGACA           | 5581                |      |
| Qy | 1861 | SerValThrPheGlu    | CysLeuProGlyTyr    | GlnLeuThrGlyHisProValLeuThrCys          | 1881                |      |
| Db | 5581 | TCAGTGACCTTGA      | TGCTGCCGGGATCAT    | ATGACATGGCACTGCTCTCCACGCT               | 5641                |      |
| Qy | 1881 | GlnHisGlyThr       | AsnArgAsnTrpAsnHis | ProLeuProLysCysGluValProCysGly          | 1901                |      |
| Db | 5641 | CACATGGGACCA       | CCGGAACCTGGAGCA    | CCCCCTGCCAATGTGAAGTCCCTTGTGGC           | 5701                |      |
| Qy | 1901 | GlyAsnIleThr       | SerSerAsnGlyThr    | ValTyrSerProGlyPheProSerProTyrSer       | 1921                |      |
| Db | 5701 | GGGAACATCATCT      | TTTCCACGCGCAT      | GTGTACTCCCGGGGTTCCCTAGGCCCGTACTCC       | 5761                |      |
| Qy | 1921 | SerSerGlnAspCys    | ValTrpLeuIleThrVal | ProIleGlyHisGlyValArgLeuAsn             | 1941                |      |
| Db | 5761 | AGCTCCAGAGACT      | GTGTGGCTGATATCA    | CCGTGCCCAATGGCCATGGGCTCGGCTCAAC         | 5821                |      |
| Qy | 1941 | IleuSerLeuLeuGln   | ThrGlnProSerGly    | AspPheIleTrpIleTyrAspGlyProGln          | 1961                |      |
| Db | 5821 | CTCAGCTGTGAC       | GACAGACAGACCCCT    | CTGGAGATTTTCAACACATCTGGAGGGGCCACAG      | 5881                |      |
| Qy | 1961 | GlnThrAlaPro       | ArgLeuGlyValPhe    | ThrArgSerMetAlaLysThrValGlnSer          | 1981                |      |
| Db | 5881 | CAAAAGACCA         | CCGCGCTCGGCGTCTT   | CAACCCGAGACATGGCCAAAGAAACAGTCCAGAGT     | 5941                |      |
| Qy | 1981 | SerSerAsnGlnVal    | IleLeuLeuLysPhe    | HisArgAspAlaAlaThrGlyGlyIlePheAla       | 2001                |      |
| Db | 5941 | TCATCCACCA         | CGGCTCTGCTCA       | AGTTCCACCGTATGACGCCACAGGGGGGATCTTGCC    | 6001                |      |
| Qy | 2001 | IleAlaPheSerAla    | TyrProLeuThrLys    | CysBProProThrIleLeuProAsnAla            | 2021                |      |
| Db | 6001 | ATAGCTTCTCCG       | CTTATCCATCCAA      | CAATAGCCCTCTCCACATCTCTCCCAAGCC          | 6061                |      |
| Qy | 2021 | GluValValThr       | GlnAsnGlnGlu       | PheAsnIleGlyAspIleValAlaGlyTyrArgCysLeu | 2041                |      |
| Db | 6061 | GAACTCGTCA         | CAGAAATGAAGAA      | TTCAATATAGATAACATCGTACCGCTACAGATGCTTC   | 6121                |      |
| Qy | 2041 | ProGlyPheThrLeu    | ValGlyAsnGlnIle    | LeuThrCysLysLeuGlyThrTyrLeuGln          | 2061                |      |
| Db | 6121 | CCTGGCTTACCT       | TAGTGGGAAATGA      | AATCTGACCTGCAACCTTGAACTTACCTGCTGAG      | 6181                |      |
| Qy | 2061 | PheGlnGlyPro       | ProProTyrCysGlu    | ValHisCysBProThrAsnGluLeuLeuThrAsp      | 2081                |      |
| Db | 6181 | TTTAAAGACCA        | CCCCCGAATATG       | GAAGTCACTGTCCAA                         | CAATGAGCTTCTGACAGAC | 6241 |
| Qy | 2081 | SerThrGlyValIle    | LeuSerGlnSerTyr    | ProGlySerTyrProGlnPheGlnThrCys          | 2101                |      |

|    |      |   |      |
|----|------|---|------|
| Db | 6241 | TCACAGGCGTGAATCTTGAGCCAGAGCTACCCCTGGAGACTATCCCACTTCAGACCTGC           | 6300 |
| Qy | 2101 | SeTrpLeuValArgValAGluProAspPyrAsnIleSerLeuThrValGluTyrPheLeu          | 2120 |
| Db | 6301 | TCCTGGCTGGAGAGTGGAGCCGACTATACATCTCCACAGAGAGTACTCTCTC                  | 6360 |
| Qy | 2121 | SerGluYsgInTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeu           | 2140 |
| Db | 6361 | AGCCAGAGCAATATGATGAGTTTGATGATTTTATGATGATCCATCAGACAGAGTCTCTG           | 6420 |
| Qy | 2141 | LeuIysAlaLeuSerGlyAsnTyrSerAlaProLeuIleValIthrSerSerAsnSer            | 2160 |
| Db | 6421 | CTGAAGGCGCTCAGTGGGAATTAATCACTCCCTGATTTGCACAGCTCAAGCAACTCT             | 6480 |
| Qy | 2161 | ValTyrLeuMetArgTyrPseSerSerAspHisAlaTyrAsnAlaGlyPheIleLeuGlyIleArgTyr | 2180 |
| Db | 6481 | GTGATACCTGGCTGATCTGATCTCAGCCATCAATCGAAGGGGCTTCMAATCCGCTAT             | 6540 |
| Qy | 2181 | SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyIleThr          | 2200 |
| Db | 6541 | TCAGCCCTTACTGAGCTGCGCCAGGCTCCACTCCATGCTTCATCTTAGCCAGACAC              | 6600 |
| Qy | 2201 | SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly          | 2220 |
| Db | 6601 | AGCAACCCAGCCCGGGGCTCCATCCACTTGGCTGCAACGCGCGCTACCGCTGTGGGA             | 6660 |
| Qy | 2221 | HisSerMetAlaIleCysThrArgHisAspGlnGlyTyrHisLeuTyrPseSerGluAlaIle       | 2240 |
| Db | 6661 | CACAGCATGGCCATCTGTACCCGCGACCCCAAGGCTTACCACTGTGGAGGAAGCCATC            | 6720 |
| Qy | 2241 | ProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProIysAsnGlyMetValPhe          | 2260 |
| Db | 6721 | CTCTCTGTCAAGCTCTTCTCTGGGCTTCTCTAGGCCCCCAAGATGAGATGCTGTTT              | 6780 |
| Qy | 2261 | GlyYsgInTyrThrValGlyIleThrYsaAlaValTyrSerCysSerGlyGlyTyrHisLeu        | 2280 |
| Db | 6781 | GGCAAGAGGTACACAGTGGGAACCAAGGCGGTGACAGCTGAGTGAAGGCTACCACTC             | 6840 |
| Qy | 2281 | GlnAlaGlyAlaGluAlaThrAlaGluCysAlaLeuAspThrGlyLeuTyrPseSerAsnArgAsn    | 2300 |
| Db | 6841 | CAGCAGGCGGTGAGGCCACGACAGAGTGTGACACAGGCTTATGAGAGCAACCGCAT              | 6900 |
| Qy | 2301 | ValProProGlnCysValProValIthrCysAspAspValSerSerIleSerValGlnHis         | 2320 |
| Db | 6901 | GTCCACCAACAGTGTCTCTGTAATCTTCTGATGTCAATAGCATCAGGCTGAGAT                | 6960 |
| Qy | 2321 | GlyArgTyrArgLeuIlePheGluIthrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIle         | 2340 |
| Db | 6961 | GGCCGATGGAGGCTTATCTTTGAGACACAGTATCAGTTCCAGGCCAGCTGATCTCATC            | 7020 |
| Qy | 2341 | CysAspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyIle          | 2360 |
| Db | 7021 | TGTATACCTGTGCTACTACTATACGCCAAAGGTCATCCGCTCAGGCCAAATGGCAAA             | 7080 |
| Qy | 2361 | TyrPseLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeuProIle          | 2380 |
| Db | 7081 | TGAGACCTCGGAGACTTACGCCCACTGCGGAATCATCTCTGGAGAGACTCCGAT                | 7140 |
| Qy | 2381 | ProProAsnGlnIleArgIleGlyIleThrLeuSerValTyrGlyAlaIleAlaIlePheSer       | 2400 |
| Db | 7141 | CCCCCAATGGCCACCGCATCGGAACATGTCTGTCTACCGGGGCAACAGCATTTCTCC             | 7200 |
| Qy | 2401 | CysAsnSerGlyTyrThrLeuValGlySerArgValArgGluCysMetAlaAsnGlyLeu          | 2420 |
| Db | 7201 | TGCATTTCCGATACACACTGTGGGCTCCAGGGGCTGTGATCATGCGCAATGGGCTC              | 7260 |
| Qy | 2421 | TyrPseGlySerGlyValArgCysLeuAlaGlyIleHisCysGlyIleThrProGluProIle       | 2440 |
| Db | 7261 | TGAGTGGCTGTAGATCCGCTGCTTGTGTGACATGTGGGACTCTGAGCCCAATGTC               | 7320 |
| Qy | 2441 | AsnGlyHisIleAsnGlyGluAsnTyrSerTyrArgGlySerValIleTyrGlnCysAsn          | 2460 |
| Db | 7321 | AACGAGACATCATGGGGAGAACATACAGCTACCGGGGCACTGTGTGATCAATGCAAT             | 7380 |
| Qy | 2461 | AlaGlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnIleAspHisIleTyrPse       | 2480 |
| Db | 7381 | GCTGGCTCCGCTGATGGAGATGTCTGTGCGATCTGCGACAGGAGATCATCTGATGTC             | 7440 |
| Qy | 2481 | GlyYsgInProPheCysValProIleThrCysGlyHisAspGlyAsnProValAsnGly           | 2500 |
| Db | 7441 | GGCAGACCCCTTCTGTGTGCCAATTAATCTGTGGACACCCAGGCAACCTGTCAAGCC             | 7500 |
| Qy | 2501 | LeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValIleYsaPheValCysAsnProGly          | 2520 |
| Db | 7501 | CTCAGTCAAGGATACCACTTTAACTCAAGATGTGTGCAAGTTTGTTCACACCTCGGG             | 7560 |
| Qy | 2521 | TyrMetAlaGlnGlyAlaAlaArgSerGlyCysLeuAlaSerGlyGlnTyrPseSerAspMet       | 2540 |
| Db | 7561 | TATATGGCTAGGGGCTGTAGGTCCCAATATGCTGCGCCAGGGGCAATGAGTACATG              | 7620 |
| Qy | 2541 | LeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGlnIleAsnSerValArg       | 2560 |
| Db | 7621 | CTGCCCACTGGAGAAATCATCACTGTACAGATCTGTGACACCAAGAAATAGTTCGT              | 7680 |
| Qy | 2561 | GlnValHisAlaSerGlyProHisArgPheSerPheGlyIleThrValSerTyrArgCys          | 2580 |
| Db | 7681 | CAGGTCCACCCAGCGGCGCCAGAGGTTCAAGTTCCGACCACTGTGTCTTACCGGTGC             | 7740 |
| Qy | 2581 | AsnHisGlyPheTyrLeuLeuGlyIleThrProValLeuSerCysGlnGlyAspGlyIleThr       | 2600 |
| Db | 7741 | AACCAAGGCTTCTTACCTCTCGGGACCCCAAGTGTCAAGCTGCGCAAGGAGATGACATGG          | 7800 |
| Qy | 2601 | AspArgProArgProGlnCysLeuLeuValSerCysGlyHisAspGlySerProProHis          | 2620 |
| Db | 7801 | GACGCTCCCGCCCGCCAGTGTCTTGTGTCTCTGTGTCATCCGGCTCCCGCTCAC                | 7860 |
| Qy | 2621 | SerGlnMetSerGlyAspSerTyrThrValGlyAlaValArgTyrSerCysIleGly             | 2640 |
| Db | 7861 | TCCAGATGTCTGAGACAGTTTATCTGTGGAGAGTGTGGGTGACACTGCATCGGCG               | 7920 |
| Qy | 2641 | LysArgThrLeuValGlyAsnSerThrArgMetCysGlyIleAspGlyHisIleTyrThrGly       | 2660 |
| Db | 7921 | AAGGTATCTGTGGGAAACAGACCCGATGTGTGGGTGATGACACTGACACTGGC                 | 7980 |
| Qy | 2661 | SerLeuProHisCysSerGlyIleThrSerValGlyValCysGlyAspProGlyIleProAla       | 2680 |
| Db | 7981 | TCCTCCCTCACTGTCTAGAAACCAAGCTGGAGTTGGCGTGAACCTGGGATCCCGCT              | 8040 |
| Qy | 2681 | HisGlyIleArgLeuGlyAspSerPheAspProGlyIleThrValIleArgPheSerCysGlu       | 2700 |
| Db | 8041 | CATGGCATCGTTTGGGGACACCTTGTATCCAGGCACTGTGATGCGCTTCAGCTGGA              | 8100 |
| Qy | 2701 | AlaGlyHisValLeuMetArgGlySerSerGluArgThrCysGlnAlaAsnGlySerTyrPse       | 2720 |
| Db | 8101 | GCTGGCCACGTGTCCCGGAGTGTCTAGACGCACTGTGAACCAATAGCTGTGGAGC               | 8160 |
| Qy | 2721 | GlySerGlnProGlnCysGlyValIleSerCysGlyAsnProGlyIleThrProSerAsnAla       | 2740 |
| Db | 8161 | GGCTCGAGCTGAGTGTGAGATCTTGTGGGAACCTGGGACTCCAAGTATGCC                   | 8220 |
| Qy | 2741 | ArgValAlaPheSerAspGlyLeuValPheSerSerIleValTyrGluCysArgGlu             | 2760 |
| Db | 8221 | CGAGTGTGTGATGATGCTGCTGTTTCTCCAGCTCTATCGTATGATGATGCTCGGGA              | 8280 |
| Qy | 2761 | GlyTyrTyrAlaThrGlyLeuLeuSerArgHisCysSerValAsnGlyIleThrTyrThrGly       | 2780 |
| Db | 8281 | GGATACACCCCAACAGGCTGTCAAGCGCTGCTCGGTCAATGTGTACTGGACAGGC               | 8340 |
| Qy | 2781 | SerAspProGlnCysLeuValIleAsnCysGlyAspProGlyIleProAlaAsnGlyLeu          | 2800 |
| Db | 8341 | AGTGAACCTTAGTGCTCTGATTAATCTGTGTGACCTTGGAATTCAGCCATAGCTT               | 8400 |
| Qy | 2801 | ArgLeuGlyAsnAspPheArgTyrAsnLysThrValIleThrTyrGlnCysValProGlyTyr       | 2820 |
| Db | 8401 | CGGCTGGCAATGACTTCAGGTACAAACAAATGTGTACATATCAGTGTGCTCCGTGAT             | 8460 |

```

QY 2821 MetMetGluSerHisArgValSerValLeuSerCysThrIleuAspArgThrTyrAsnGly 2840
DB 8461 ATGATGAGAGTCACTAGAGTATCTGTCTGAGCTGCACCAAGAGACGGACATGGAAATGGA 8520
QY 2841 ThrIleuProValCysValAlaLeuMetCysIleuProProleuIleProAsnGlyLys 2860
DB 8521 ACCAAGCCCGCTCTGCAAGAGCTCTCATGTGCAAGCACCTCCGCTCATCCCAATGGGAG 8580
QY 2861 ValValIleuSerAspPheMetTyrGlySerSerValThrTyrAlaCysLeuIleuGlyTyr 2880
DB 8581 GTGTGTGGGCTGACTTATATGTGGGGCTCAAGTGTGACTTATGCTCTGCTGGAGGGGTAC 8640
QY 2881 GluLeuSerLeuProAlaValPheThrCysGluIleuAsnGlySerTyrPheGlyLeu 2900
DB 8641 CAGCTCTCTCCGCGGGGCTTTCACCTGTGAGGAAATGGTCTCTGACCGGAGAGACTG 8700
QY 2901 ProGlnCysPheProValPheCysGlyAspProGlyValProSerArgIleuArgGly 2920
DB 8701 CCTAGTGTCTCCCTGTGTCTGTGGGGATCTGTGTCCGCTCCGCTGGAGAGAGAG 8760
QY 2921 AspArgGlyPheSerTyrArgSerSerValSerPheSerCysHisProProleuValLeu 2940
DB 8761 GACCGAGGCTTCTCTACAGCTCATCTGTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTG 8820
QY 2941 ValGlySerProArgArgPheCysGlnSerAspGlyThrTyrSerGlyThrGlnProSer 2960
DB 8821 GTGGGCTCTCCACCAAGTTTGGCAGTGTGAGCATGGAGTGGCACCCAGCCAGCCAGC 8880
QY 2961 CysIleuAspProThrIleuThrCysAlaAspProGlyValProGlnPheGlyIleu 2980
DB 8881 TGCATTAATCCGACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 8940
QY 2981 AsnAsnSerGlnGlyTyrGlnValIleuSerThrValLeuPheArgCysGlnIleuSerGly 3000
DB 8941 AACAATTCTCAGGCTACAGAGTTGGAACAGACAGCTCTCTCTGCTGCTGCTGCTGCTG 9000
QY 3001 LeuLeuGlnGlySerThrThrArgThrCysLeuProAsnLeuThrTyrSerGlyThrPro 3020
DB 9001 CTGCTTACAGGCTCTCCACCAAGGACCTGCTCCCAAACTGACCTGAGTGGAAACCCCA 9060
QY 3021 ProAspCysValProHisHisCysArgGlnProGlnThrProThrHisAlaAsnValGly 3040
DB 9061 CCGACCTGTGCTCCCAACCACTGACGACGACGACGACGACGACGACGACGACGACGACG 9120
QY 3041 AlaLeuAspLeuProSerMetGlyTyrThrLeuIleThrProAlaArgArgAlaSerPro 3060
DB 9121 GCCCTGATTTGGCCTTCATGGGCTACAGCTCATTACTCTCTCCAGAGGAGCTTCTCC 9180
QY 3061 SerArgValAlaProSerThrAlaProAlaArgArgMetAlaAlaGlyGlnAlaSerArg 3080
DB 9181 TCAAGGCTGCTCCGAGCACTGCACTGCAAGGCTGCAAGTGGCACTGCAAGGCAAGCCGC 9240
QY 3081 ProSerAlaTyrArgSerGlyProValGlyAspProSerThrLeuProGlySerHisArg 3100
DB 9241 CCACTGTGCTGAGAGTCCGAGCCCAAGTGGAGAGCACTCAACACTGCTCCGAGAGCCAGC 9300
QY 3101 SerProIleuPro 3144
DB 9301 TCACCAAGCCT 9312

```

## RESULT 2

```

US-10-016-248-3
; Sequence 3, Application US/10016248
; Publication No. US20040033491A1
; GENERAL INFORMATION:
; APPLICANT: Alsebrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-218
; CURRENT APPLICATION NUMBER: US/10/016,248
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/254,329
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/291,037

```

```

; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/255,648
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/297,173
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: 60/309,258
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/326,393
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/315,639
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 8010
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-248-3

```

## Alignment Scores:

```

Pred. No.: 0 Length: 8010
Score: 14142.50 Matches: 2605
Percent Similarity: 97.71% Conservative: 1
Best Local Similarity: 97.68% Mismatches: 2
Query Match: 83.26% Indels: 59
DB: 16 Gaps: 3

```

US-10-016-248-2 (1-3104) x US-10-016-248-3 (1-8010)

```

QY 1 MetAlaGlyAlaProProProAlaLeuLeuProCysSerLeuIleSerAspCys 20
DB 1 ATGGGGGGGGCCCCCCCCCCCCCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
QY 21 AlaSerAsnGlnArgHisSerValGlyIleProSerGlyLeuValIleLysGlnIle 40
DB 61 GCTACCAATCAGCCGACCTCCGTTGGGCTAGACCTCCGAGCTGATCAAGAACCAATT 120
QY 41 GluLeuLysSerArgGlyValIleLeuMetProSerLysAspAsnSerGlnLysThrSer 60
DB 121 GAGTTGAGAGTCTCGAGGTGATGAGTATGCCAGCAAGCAACAGCAGCAAGCCTCT 180
QY 61 ValLeuThrGlnValGlyValSerGlnGlyHisAsnMetCysProAspProGlyIlePro 80
DB 181 GGTTAATCTAGGTTGGTGTGTCCCAAGAGCATATAATGTGTCCAGACCTGGCATACC 240
QY 81 GluArgGlyLysArgLeuGlySerAspPheArgLeuGlySerSerValGlnPheThrCys 100
DB 241 GAAGAGGGCAAAAGACTAGGCTCGGATTTCAGTTAGGATCCAGCTCCAGTTCACTGC 300
QY 101 AsnGluGlyTyrAspLeuGlnGlySerLysArgIleThrCysMetLysValSerAspMet 120
DB 301 AACAGGGCTTATGACTGCTCAGAGGCTCAGAGGCTCAGAGGCTCAGAGGCTCAGAGGCT 360
QY 121 PheAlaAlaTyrSerAspHisArgProValCysArgAlaArgMetCysAspAlaHisLeu 140
DB 361 TTTGGGGCTCGAGAGCAACAGGCAAGTGTCCGAGGCGGCAATGTGTGAGGCCACTT 420
QY 141 ArgGlyProSerGlyIleIleThrSerProAsnPheProIleGlnTyrAspAsnAla 160
DB 421 CGAGGCCCTCGGCAATCATCACTCCCAATTTCCTCATGTATGACAAATGCA 480
QY 161 HisCysValTyrIleIleThrAlaLeuAsnProSerLysValIleLysLeuAlaPheGly 180
DB 481 CACTGTGTGTGATCATCAAGCACTCAACCTCCCAAGTATCAAGCTGCTGCTTGGAG 540
QY 181 GluPheAspLeuGluArgGlyTyrAspThrLeuThrValGlyAspGlyGlnAspGly 200
DB 541 GAGTTGATTTGGAGAGGGGCTATGACACCTGACGCTGATGTGTGTGAGAGAGGG 600
QY 201 AspGlnLysThrValIleuTyrMetSerGlnAsnAlaCysSerAspSerProHisThrPro 220
DB 601 GACCAAGAAAGACATTCTCATATGTCTCAAAATGCTCGAGTACAGACCTTCACACCCA 660

```

[illegible]

|    |      |  |      |
|----|------|--|------|
| Ds | 1741 | TACCAAGGAAACCAAGGTTCCCAAGTTCATATGACACCAAGCACTACCTTACCTCCTC         | 1800 |
| Qy | 601  | PheSerThrIAspIysSerHisSerAspIleGlyPheGlnLeuArgTyrGluThrIleThr      | 620  |
| Ds | 1801 | TTCTCTACCGACAAGAGTCACTCGACATCGGCTTCACAGTCCGCTATGACTATTAACA         | 1860 |
| Qy | 621  | LeuGlnSerAspHisSerCysLeuAspProGlyIleProValAsnGlyGlnArgHisGlyAsn    | 640  |
| Ds | 1861 | CTGCAAGTCACAGCACTGCTGTGATCCAGAGATCCCGATTAAATGAGCAGGTCATGGGAAT      | 1920 |
| Qy | 641  | AspPheTyrValGlyAlaLeuValIleThrPheSerCysAspSerGlyTyrThrIleuSerAsp   | 660  |
| Ds | 1921 | GACTTCTACGGGGCGCGCTGGAGACCTTACGCTGTGACTCGGGCTACACATTAAAGTGAC       | 1980 |
| Qy | 661  | GlyGluProLeuGlnGlyCysGluProAsnPheGlnIleTyrSerArgAlaLeuProSerCysGlu | 680  |
| Ds | 1981 | GGGAGAGCTCTCGAGAGTGTAGGCCCACTTCAGTGAAGCGGGCCCTGCCAGTTGTGA          | 2040 |
| Qy | 681  | AlaLeuCysGlyGlyPheIleGlnGlySerSerGlyTyrIleuSerProGlyPhePro         | 700  |
| Ds | 2041 | GCTCTCTGTGTGGCTTCATTCAAGGCTCCAGTGGAGACCATCTTTCGCGACAGGTTCCCT       | 2100 |
| Qy | 701  | AspPheTyrProAsnLeuAsnLeuAsnGlySerThrIleIleGluThrSerHisGlyIysGly    | 720  |
| Ds | 2101 | GACTTCTACCCCAACAACCTGAACCTGACCTGGAATTATGAACAACCTCATGGCAAGGAT       | 2160 |
| Qy | 721  | ValPhePheThrPheHisThrPheHisLeuGlnSerGlyHisAspTyrLeuLeuIleThr       | 740  |
| Ds | 2161 | GTCGTTCTTCACTTTCCACACCTTCCACCTGGAAAGTGGCATGACACTCTCTCATCACT        | 2220 |
| Qy | 741  | GluAsnGlySerPheThrGlnProLeuArgIleLeuThrGlySerArgLeuProAlaPro       | 760  |
| Ds | 2221 | GAGAAACGGACAGTTCAACCCAGCCCTGAGGCACTGAACCTGATCTCGCTGCCAGCTCC        | 2280 |
| Qy | 761  | IleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer       | 780  |
| Ds | 2281 | ATCAGCGCTGGGCTCTATGGCAACTTCACGCCAGGTCGCTTCACTCTGATTTCTCC           | 2340 |
| Qy | 781  | MetSerTyrGluGlyPheAsnIleThrPheSerGlyTyrAspLeuGluProCysGluGln       | 800  |
| Ds | 2341 | ATGTCATATAGAGATTCAACATCACCTTCTCAGAGTAGACTTGAAGCCCTGTAGAGAG         | 2400 |
| Qy | 801  | ProGluValProAlaTyrSerIleArgIysGlyLeuGlnPheGlyValIleAspThrLeu       | 820  |
| Ds | 2401 | CCGAGAGTCCAGGCTTACAGCATCCGGAAGGGCTTGCAATTGGCTGGGGCGACACTTG         | 2460 |
| Qy | 821  | ThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIleThrCysLeuGly       | 840  |
| Ds | 2461 | ACCTTCTTCCTGCTTCCCGGGGTACGCTGTGAGAGGACCGCCGCACTACGTCGCGGG          | 2520 |
| Qy | 841  | GlyArgArgArgLeuThrTyrSerSerProLeuProArgCysValAlaGluCysGlyAsnSer    | 860  |
| Ds | 2521 | GCGACAGCGGCGCTGTGAGACTGCGCTTCCCAAGGTGTGTTCGTAGTGTGGAAATTCA         | 2580 |
| Qy | 861  | ValIleThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTyrAsnAsnAsn    | 880  |
| Ds | 2581 | GTCACAGGCACTCAGGAGTACTTTGCTGTGTCGCCCACTTTCGTGTGAATCAATTAACAT       | 2640 |
| Qy | 881  | HisGluCysIleTyrSerIleGlnThrGlnProGlyIysGlyIleGlnLeuIysAlaArg       | 900  |
| Ds | 2641 | CATCAAAAGCATCTACCTCCATCCAGACCCAGCCAGGGAGAGGAATTACGCTGAAGCAG        | 2700 |
| Qy | 901  | AlaPheGlyLeuSerGlyIysAspValIleLeuValTyrAspGlyAsnAsnAsnSerAla       | 920  |
| Ds | 2701 | GCATTTCACACTCTCCGAAGAGATGTCTTCAGGTTTATGATGGCAACAACATTCGCC          | 2760 |
| Qy | 921  | ArgLeuLeuGlyValPheSerHisSerGlyMetMetGlyValIleThrLeuAsnSerThrSer    | 940  |
| Ds | 2761 | CGTTTGTGGGAGTTTATTACCATTCGAGATGATGGGGGTGACTTTGAACAGACATCC          | 2820 |
| Qy | 941  | SerSerLeuThrLeuAspPheIleThrAspAlaGluAsnThrSerTyrGlyPheGluLeu       | 960  |

|    |      |   |      |
|----|------|---|------|
| Db | 2821 | AGCAGCTGTGGCTTGATTTTCACTCATGCTGTAAGAAACACAGCAAGGGCTTTGAACTG       | 2880 |
| Qy | 961  | HhApheSerSerPheGluLeuIleuYsCySgiuAspProGlyThrProIysPheGlyTyr      | 980  |
| Db | 2881 | CACCTTTTCCAGGCTTTGAACTCATCAAAATGTGAGGACCAGGAACCCCAAGTTGGCTAC      | 2940 |
| Qy | 981  | LysValHisAspGluGlyHisIshPheHisIagiYsSerSerValSerPheSerCyAspProGly | 1000 |
| Db | 2941 | AAGGTTCAATGATGAAAGTATTTTGTGAGGAAGCTCGTGTCTTAAGCTGTGACCTCGGA       | 3000 |
| Qy | 1001 | TyrSerLeuAspGlySerGluGluLeuLeuCySLeuSerGlyGluuAspArgThrTyrAsp     | 1020 |
| Db | 3001 | TACAGCTTCGGGGGTATGATGAGAGAGCTGCTGTCTGAGTGGAGAGCGCCGAGCTGGGAC      | 3060 |
| Qy | 1021 | AspProLeuProThrCysValHisIagiLysGlyYglYThrValArgGlyGluValSerGly    | 1040 |
| Db | 3061 | CGAGCTCTGACCACTGTGTCCCGCAGTGTGGAGGAGACAGTGAAGGAGAGTGTCTGGGG       | 3120 |
| Qy | 1041 | GluValLeuSerProGlyTyrProAlaProYArgLuhHisLeuAsnCySLeTyrThr         | 1060 |
| Db | 3121 | CAGGTGCTGTCAACCGGGATCTGACCTCCCTATGAACATCTCAATCTGATCTGGACC         | 3180 |
| Qy | 1061 | IleGluHisIagiLuhHisGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGluGlu  | 1080 |
| Db | 3181 | ATCGAAGCAGAGCGCGCGCTGCACCAATTGGCTACACTTCCGAGTGTGGACACAGAGAG       | 3240 |
| Qy | 1081 | ValHisAspValLeuAspArgIleTyrAspGlyProValGluSerGlyValLeuLeuYsGlu    | 1100 |
| Db | 3241 | GTTCAAGACGTCTCTCGCATCTGGGATGGGCTGTGGAGAGCGGGGTTCTGTAAAGAG         | 3300 |
| Qy | 1101 | LeuSerGlyProAlaLeuProIysAspLeuHisSerThrPheAsnSerValValLeuGln      | 1120 |
| Db | 3301 | CTGAATGGCCCGCCCTCGCTCCCAAGAGCCTGCAATGACACTTCACTCGATGCTCTCAG       | 3360 |
| Qy | 1121 | PheSerThrAspPhePheThrSerTyrGlnGlyPheAlaIleGlnPheSerValSerThr      | 1140 |
| Db | 3361 | TTGACGACTGACTTCTTCAACCAAGCAGAGGAGCTTGCATTCAAATTTTCAGTGTCCACA      | 3420 |
| Qy | 1141 | AlaThrSerCyAsnAspProGlyIleProGluAsnGlySerArgSerGlyAspSerTyr       | 1160 |
| Db | 3421 | GCAAGTCTCGCAATGACCCTGGAGTCCCGCAGATGGGAGTCCGAGTGGTGAACGTTGG        | 3480 |
| Qy | 1161 | GluHisIagiYsAspSerThrValPheGlnCyAspProGlyTyrAlaLeuGlnGlySerAla    | 1180 |
| Db | 3481 | GAAAGCGGAGACTCCACAGTGTCTCCAGGTACCTGTGGCTAAGGTCGACGAGGAAGTGA       | 3540 |
| Qy | 1181 | GluIleSerCySValIysIleGluAsnArgPhePheTyrGlnProSerProProThrCys      | 1200 |
| Db | 3541 | GAGATCAGCTGTGTGAATGAGAGAACAGTTCCTTCTGGCAGGCCCAACCCGCAACATGC       | 3600 |
| Qy | 1201 | IleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr      | 1220 |
| Db | 3601 | ATGCTCTCCCTCGGGGAGACTTCAACAGACCACTGGAGTCACTCTCAACCAAAATTAC        | 3660 |
| Qy | 1221 | ProGluProTyrProProGlyLysGluCysAspTrrIysValThrValSerProAspTyr      | 1240 |
| Db | 3661 | CCAGAACCCTACCCGCCAGGACGAGAGAGTGTACTGGAAAGTACCCTTCAACCAACTAC       | 3720 |
| Qy | 1241 | ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle      | 1260 |
| Db | 3721 | GTCATCGCCCTCGATATTAAATCATCTTTAACTGAGAGCTGGCTGTGACTTCCTCCATATC     | 3780 |
| Qy | 1261 | TyrAspGlyAspAspSerLeuSerProLeuIleGlySerPheTyrGlyYsSerGlnLeuPro    | 1280 |
| Db | 3781 | TACAGCGAGCGGACTCTCTAGCCCTTCTATAGAAAGCTTCAATGAGCTCCAGCTCCCA        | 3840 |
| Qy | 1281 | GlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal      | 1300 |
| Db | 3841 | GGCCGCATTGAAGAGAGCAGCAACAGCTCTTCCGCTTCCGACAGCATGATCTGTG           | 3900 |
| Qy | 1301 | SerAsnHisIagiPheValIleAspTyrThrGluAsnProArgGluSerCySAspPro        | 1320 |
| Db | 3901 | AGCAATGTGGCTTGTGATTTGACTTAAACGAAACCCCGGGAGTCAATGTTTTGTATCTC       | 3960 |

|    |      |   |      |
|----|------|---|------|
| QY | 1321 | GIYSEIILEYSAAGCIYTHRAYGAIYGERAPLEUYISLEUGIYSESERVALTHR          | 1340 |
| DB | 3961 | GGTTCATCATMAAGAGGACACGGGTGGGGTCCAGCTCGAAGCTGGGCTCTCCGTCAAC      | 4020 |
| QY | 1341 | YTYRYCYEHAEGIYGIYTYRGIVUAGIUGIYTHRSERTHLEUSERCYEILEUGIY         | 1360 |
| DB | 4021 | TACTACTGGCACGGGGGCTACGAGATTGAGGGCACTCGACCTCGAGCTGCATCTGGGG      | 4080 |
| QY | 1361 | PROAARGIYLYSPROVALITTPASNAENPROARPROVALCYETHRALAPROCYGIGIYGIY   | 1380 |
| DB | 4081 | CCGTATGGGAAGCCCGGTGGAGCAATCCCGGCGAGCTGCACAGCCCCCTGGGGGGA        | 4140 |
| QY | 1381 | GIINTYVALIGIYSESAERAGIYVALIValLEUSERPROANTYRPROGINANTYRTHSER    | 1400 |
| DB | 4141 | CAGTATGGGGGTGGGACGAGAGGTCTTGTCCCCCAACTACCCCGAGACTACACAGT        | 4200 |
| QY | 1401 | GIYGINIIIECYELEUYRPHETHRALPROLYSAPTYRVALIAPHEGIYGINIPE          | 1420 |
| DB | 4201 | GGACAGACTGCTGTTATTTGTTACTAGTGGCCAGAGACTATGTGTGTGGTGGCAGTTC      | 4260 |
| QY | 1421 | AIAPHEPHEIATHRALALEUASAPVALIQUVAIHISAPGYHISERGINHIS             | 1440 |
| DB | 4261 | GCCTCTTTCACAGGCCCCCTACAGACGTGTGAGGTTCACGACGGCCACAGCCACAC        | 4320 |
| QY | 1441 | SEARATLEULEUSERSESERLEUSERGISETHIETHRGISLEUSERPOLEUALATHR       | 1460 |
| DB | 4321 | TGGCGGCTCTTCAGAGCTCTCTCGGGGTCCTCATCAGAGATCATCGCTTCGGCCAC        | 4380 |
| QY | 1461 | SEERANGINVALILEUIILEYSPHESERIALAYGELYLEUALAPROAIAPAGGIYPHEHIS   | 1480 |
| DB | 4381 | TCCAAATCAAGTTCATTAAAGTTAGGCGCAAGAGCCCTCGCACAGCCAGAGGGCTTCAC     | 4440 |
| QY | 1481 | PHETHVALTYRGINALVALIPROARGTHSERIALATHRGINCYSSERSESERVALI        | 1500 |
| DB | 4441 | TTGTGCTACCAAGCGGTTCTCTCGAACAGCGCCACGAGTGCAGCTCTGTGCCGGAACC      | 4500 |
| QY | 1501 | ARGTYRGIIYLYBARGLEUGIYSEASPHESERVALIYALILEVALIARCPHEGINUCYS     | 1520 |
| DB | 4501 | CGCTATGCGCAAGAGCGCTGGCGAGCACTTCGCGTGGGGCCATCGTCCGCTTCGAATGC     | 4560 |
| QY | 1521 | AMNSERGIYTYRALALEUGINGIYSEPROGINIILEIUCYSELEUPROVALI            | 1540 |
| DB | 4561 | AACTCCGGCTATGCCCTCGACAGGGGTCCCAAGATTCAGATGCCCTCTGTGGCTGGGGCC    | 4620 |
| QY | 1541 | LEUALAGINTYPAENVALISERIALPROTHRCYSAVALIAPROCYSGIYGIYASMLEUTHR   | 1560 |
| DB | 4621 | TTGGGCCCATGGAATGCTTCAGGGCCAGCGTGTGTGGTCCGTGTGGAGCCAACTTCACA     | 4680 |
| QY | 1561 | GIUATRGATAGGIYTHRIILEUSERPROGIYPHEPROGIUPROTYRLEUANSEERLEUAN    | 1580 |
| DB | 4681 | GAGGCGAGGGGCACCAATCTGTCTGCCCTGGGTTCTCCCAAGCCGTAACCTCAACACCTCAAC | 4740 |
| QY | 1581 | CYSVALITRPIYVILIEVALIAPROGIUGIYALAGIYIIEGINIIEGINIVALISERPHE    | 1600 |
| DB | 4741 | TGTGTGTGGAGATCGTGTGTCCCGAAGGGGCTGGCATTCAGATCCAAAGTGTTCAGTTTT    | 4800 |
| QY | 1601 | VALIHRGUGIINANTYRPAESPSEERLEUGIYVALIAPHEASPGIYALIASPAENTHRALTHR | 1620 |
| DB | 4801 | GTCACAGAGCGAAGACTGGGACTCGCTGGAGATTTGATGTGTCACATTAACACTGTAAAC    | 4860 |
| QY | 1621 | METLEUGIYSEPHESERGIYTHRTHVALIPROAILLEUENAMSEERTHSRERANGIN       | 1640 |
| DB | 4861 | ATGCTGGGAGATTCTACAGAACACCGTGTCCGCTTGTGAACAGCACTTCAACCAAG        | 4920 |
| QY | 1641 | LEUITYRLEUHIAPHERYRSEARAPILSESERVALISERIALALAGIYPHEHISLEUGIUTYR | 1660 |
| DB | 4921 | CTTAACTTATTTCTACTCAGATATACAGGTATTCGACGTGGCTTCCACTTGGAGTAC       | 4980 |
| QY | 1661 | LYSHTHVALIGIYLEUSERSERCYSPROGIUPROIALIAPROSERANGIYVALIYETHR     | 1680 |
| DB | 4981 | AAAAAGGTGGCCCTGACACATTTGTCCGAACTCGCTGTGGCCAGTAAACGGGTGGAAGACT   | 5040 |

Qy 1681 GYGLUAATGYrLeuValAsnAspValSerPheGlnCyGluProGlyTyrAlaLeu 1700  
| | | | |  
Db 5041 GCGAGCGCTACTGTGTAATGATGTGTCTTCCAGTGTGAGCCGGATATGCCCTC 5100  
| | | | |  
Qy 1701 GlnGlyIleAlaIleIleSerCyMetProGlyThrValArgArgTyrAsnTyrProPro 1720  
| | | | |  
Db 5101 CAGGGCCACGGCCCATCTCCGTGCATGCCGGAAACAGTGCAGCGCATGTGAATCACTCCCTC 5160  
| | | | |  
Qy 1721 ProLeuCyIleAlaGlnCyGlyGlyThrValGlnGlnMetGlnGlyValIleLeuSer 1740  
| | | | |  
Db 5161 CCACCTGTATGTGCACAGTGTGGGGAAACAGTGGAGAGATGGAGGGGTATCTGTAGC 5220  
| | | | |  
Qy 1741 ProGlyPheProGlyAsnTyrProSerAsnMetAspCySerTyrPheIleAlaLeuPro 1760  
| | | | |  
Db 5221 CCCGGCTTCCAGGGAACTACCCAGTAACATGACTCTCTCTGAAATATGCACTGCC 5280  
| | | | |  
Qy 1761 ValGlyPheGlyAlaIleIleGlnPheLeuAsnPheSerThrGlnProAsnIleAspTyr 1780  
| | | | |  
Db 5281 GTGGGCTTGGAGCTCAATCCAGTTCTTGAACTTCCACCGAGCCCAACCAAGACTAC 5340  
| | | | |  
Qy 1781 IleGlnIleArgAsnGlyProTyrGlnThrSerArgMetMetGlyArgPheSerGlySer 1800  
| | | | |  
Db 5341 ATGAAATCCGGAAATGGCCCTTATGAGCAAGCCGATGATGGAAATTCATGTGAAGC 5400  
| | | | |  
Qy 1801 GlnLeuProSerSerLeuLeuSerThrSerIleGlnThrThrValTyrPheIleSerAsp 1820  
| | | | |  
Db 5401 GAGCTTCCAAAGCTCCCTCTCTCCACGTCACAGAACACCGGTATTTCCACAGGAC 5460  
| | | | |  
Qy 1821 HisSerGlnAsnArgProGlyPheLeuLeuGlnTyrGlnAlaTyrGlnLeuGlnGly 1840  
| | | | |  
Db 5461 CACTCCCGAATCCGGCCAGGATTCAGCTGAGTTCAGGCTTATGAACTTCAAGAGTGC 5520  
| | | | |  
Qy 1841 ProAspProGlyProPheAlaAsnGlyIleValArgGlyAlaGlyTyrAsnValGlyGln 1860  
| | | | |  
Db 5521 CCAAGCCAGAGCCCTTTGCCAATGTGATGTGAGGGAGCTGGTACAAAGTGGAGCA 5580  
| | | | |  
Qy 1861 SerValThrPheGlnCyLeuProGlyTyrGlnLeuThrGlnIleAspValLeuThrCyS 1880  
| | | | |  
Db 5581 TCACTGACCTTGCAGTGTCTCCGGGGATATCAATTTGACTGGCCACCTGTCTTCACTGT 5640  
| | | | |  
Qy 1881 GlnIleGlyThrAsnArgAsnTyrAspIleAspLeuProGlyCyGlnValProCyGly 1900  
| | | | |  
Db 5641 CAACATGGACCAACCGGAACGGGACCAACCCCGCAAGTGTGAAGTCCCTTGTGCG 5700  
| | | | |  
Qy 1901 GlnAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyrSer 1920  
| | | | |  
Db 5701 GGGAAACATCACTTCTTCCAAAGGCACTGTGTATCTCCCGGGGTTCCTCAGCCGACTCC 5760  
| | | | |  
Qy 1921 SerSerGlnAspCyValTyrPheIleThrValProIleGlyIleGlyValArgLeuAsn 1940  
| | | | |  
Db 5761 AGCTCCCGAGACTGTGTGTGCTGATATCACCGTCCCATTTGGCCATGGCGCTCCCTCAC 5820  
| | | | |  
Qy 1941 LeuSerLeuLeuGlnThrGlnProSerGlyAspPheIleThrIleTyrAspGlyProGln 1960  
| | | | |  
Db 5821 CTCAGCTGTGTGACAGACAGACCCCTGTGAGATTTTCAATCCATCTGGGATGGGCCACAG 5880  
| | | | |  
Qy 1961 GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaIleValThrValGlnSer 1980  
| | | | |  
Db 5881 CAACACAGACCAACGGCTCGGGCTCTTCAACCGGAGCATGGCCAAAGAAACGTCAGAGT 5940  
| | | | |  
Qy 1981 SerSerAsnGlnValIleLeuLeuPheIleArgAspAlaIleThrGlyIlePheAla 2000  
| | | | |  
Db 5941 TCATTCACACAGCTCTGCTCAAGTTCCACCGTATGACAGCCAGGGGGATCTTGCC 6000  
| | | | |  
Qy 2001 IleAlaPheSerAlaTyrProLeuThrIleCyAspProProThrIleLeuProAsnAla 2020  
| | | | |  
Db 6001 ATACTTCTCTCGCTTATCCACTACCAAAATGCCCTCTCCACCATCTCCCAAGCC 6060  
| | | | |  
Qy 2021 GlnValValThrGlnAsnGlnGlnPheAsnIleGlyAspIleValArgTyrArgCyLeu 2040  
| | | | |  
Db 6061 GAATCTGTGCACAGAAATGAAGAAATTCATATAGTGCATCGTACCGTACAGATGCTCT 6120  
| | | | |  
Qy 2041 ProGlyPheThrLeuValGlyAsnGlnIleLeuThrCySLeuGlyThrTyrLeuGln 2060  
| | | | |

Db 6121 CCTGGCTTACTTACGTAGTGGGGAATGAATTTCTGACCTGCAAACTTGGAACTTACTGAG 6180  
| | | | |  
Qy 2061 PheGlnGlyProProPheIleCyGlnValIleCyAspProThrAsnGlnLeuLeuThrAsp 2080  
| | | | |  
Db 6181 TTTGAGGACCAACCCCGATATGTGAAGTGCATGCTCAACAAATGAGCTTCTGACAGAC 6240  
| | | | |  
Qy 2081 SerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCyS 2100  
| | | | |  
Db 6241 TCCACAGCGGTGATCTGAGCCAGAGCTTACCTGAAAGCTATCCCACTTCCAGACTTGC 6300  
| | | | |  
Qy 2101 SerThrPheValArgValGlnProAspTyrAsnIleSerLeuThrValGlnTyrPheLeu 2120  
| | | | |  
Db 6301 TCTTGGCTGTGAGAGTGGAGCCCGCATATACATCTCCCTCACAGTGGAGTACTTCTC 6360  
| | | | |  
Qy 2121 SerGlnGlyGlnTyrAspGlnPheGlnIlePheAspGlyProSerGlyGlnSerProLeu 2140  
| | | | |  
Db 6361 AGCAGAGAACCAATGTGATGATTTGAGATTTTGTATGTGCTCATCAGAGACAGAGTCTCTG 6420  
| | | | |  
Qy 2141 LeuIleValIleLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerAsnSer 2160  
| | | | |  
Db 6421 CTGAAGCCCTCAGTGGAAATTTACTCAGCTCCCTGATGTGCACACCTCAAGCAACTCT 6480  
| | | | |  
Qy 2161 ValTyrLeuArgTyrPheSerSerAspIleAlaTyrAsnArgIleGlyPheIleArgTyr 2180  
| | | | |  
Db 6481 GTGTACTGTGGTGTGCTATCTGATCAAGCTTACAAATGGAAAGGCTTCAAGATCGCTAT 6540  
| | | | |  
Qy 2181 SerAlaProTyrCySLeuLeuProArgAlaProLeuIleGlyPheIleLeuGlyGlnThr 2200  
| | | | |  
Db 6541 TCAGCCCTTATGCAACCTTGCACAGGCTTCACTCATGAGCTTCAATCCTTATGAGCAAGC 6600  
| | | | |  
Qy 2201 SerThrGlnProGlyGlySerIleIlePheGlyCyAsnAlaGlyTyrArgLeuValGly 2220  
| | | | |  
Db 6601 AGCACCAAGCCGGGGCTCTCATCTGCTGCTGCTGCAACCGCGCTACCCCTGTGTGGA 6660  
| | | | |  
Qy 2221 HisSerMetAlaIleCySThrArgIleProGlnGlyTyrIleAspLeuTyrPheSerGlnAlaIle 2240  
| | | | |  
Db 6661 CACAGCATGGCCATCTGTATCCCGGACACCCCAAGGCTTACACCTGTGAGAGCAAGCATC 6720  
| | | | |  
Qy 2241 ProLeuCyGlnAlaLeuSerCyGlyLeuProGlnAlaProIleAsnGlyMetValPhe 2260  
| | | | |  
Db 6721 CCTCTGTGTAAGCTCTTCTCGTGGGCTTCTGAGGCCCCCAAGATGGAATGTGTGT 6780  
| | | | |  
Qy 2261 GlnIleGlyThrThrValGlyThrValAlaValTyrSerCySLeuGlnGlyIleIleLeu 2280  
| | | | |  
Db 6781 GGCAGGAGTACACAGTGGAAACCAAGCCATGTACAGCTGAGTGAAGCTACACCTC 6840  
| | | | |  
Qy 2281 GlnAlaGlyAlaGlnAlaThrAlaGlnCySLeuAspThrGlyLeuTyrPheSerAsnArgAsn 2300  
| | | | |  
Db 6841 CAGGACAGCCCTGAGGCACTGCAAGTGTCTGGACACAGGCTTATGAGACCAACCGCAT 6900  
| | | | |  
Qy 2301 ValProProGlnCyVal-----ProValThrCyAspPro 2311  
| | | | |  
Db 6901 GTCCACACAGTGTGCTCGTGAAGTCTCGAGCAATGAGAGCGGGCTGTGACTGTGCT 6960  
| | | | |  
Qy 2312 AspValSerSerIleSerValGlnIleGlyValArgTyrArgLeuIlePheGlnThrGlnTyr 2331  
| | | | |  
Db 6961 GATGTCAGTGACATCAACCGTGGAGCATGGCCGATGAGGGCTTATCTTTGAGACACAGTAT 7020  
| | | | |  
Qy 2332 GlnPheGlnAlaGlnLeuMetLeuIleCyAspProGlyTyrTyrTyrThrGlyGlnArg 2351  
| | | | |  
Db 7021 CAGTTCAGGCCCAAGCTGATGCTCATCTGTGACCTTGCTATCTACTATCTGCGCAAGG 7080  
| | | | |  
Qy 2352 ValIleArgCyGlnAlaAsnGlyIleTyrPheSerLeuGlyAspSerThrProThrCyAsp 2371  
| | | | |  
Db 7081 GTCAATCCGTGTGAGGCAATGGAATGAGCTCGGGAGCTCTAGGCCCACTGCGCA 7140  
| | | | |  
Qy 2372 IleIleSerCyGlyGlyLeuProIleProProAsnGlyIleArgIleGlyThrLeuSer 2391  
| | | | |  
Db 7141 ATCAATCTCGTGGAGGCTCCGATTTCCCCCAATGGCCACCGCATGGAAACATGTCT 7200  
| | | | |  
Qy 2392 ValTyrGlyAlaThrAlaIlePheSerCyAsnSerGlyTyrThrLeuValGlySerArg 2411  
| | | | |



|    |      |  |      |
|----|------|--|------|
| Qy | 1018 | hrtPaAspArgProleuProThrCyValAlaGluCyGlyGlyThrValArgGlyGlyVal     | 1038 |
| Db | 549  | CTTGAGAGCGGCTCTGCGCACTGTGTGCGGAGTGGAGGAGCACTGAGAGAGAGG           | 608  |
| Qy | 1038 | aLSerGlyValIleuSerProGlyTyrProAlaProTyrGluHisAsnLeuAncysI        | 1058 |
| Db | 609  | TGTGGGGGCAAGTGTGTGCACTGGGATTCAGCTCCCTATGAACAATCTCAACTGCA         | 668  |
| Qy | 1058 | letPThrIleGluValaGluValaGlyCyethrIleGlyLeuHisPheLeuValPheAspT    | 1078 |
| Db | 669  | TCTGACCATGGAAGACAGAGCGGCTGACCATTTGGGCTTCACTTCTGTGTTTTACA         | 728  |
| Qy | 1078 | hrGluGluValaHisAspValIleuArgIleTPAspGlyProValGlySerGlyValIleuL   | 1098 |
| Db | 729  | CAGAGAGAGTTCAAGACAGCTGTGGGATCTGGGATGGGCTTGTGAGAGCGGGGTTCTGC      | 788  |
| Qy | 1098 | eulYsgIleuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValY      | 1118 |
| Db | 789  | TGAAGAGAGTGAATGGCCCGGCTGCGCCAGAGACCTGACATAGCACTTCAACTCGGTCG      | 848  |
| Qy | 1118 | alleuGlnPheSerThrAspPheThrSerIleGlnGlyPheAlaIleGlnPheSerV        | 1138 |
| Db | 849  | TCTGCAAGTTCAAGCACTGACTTCTTTCACACAGAGAGCGGCTTGGCAATTCATTTTCAG     | 908  |
| Qy | 1138 | aLSerThrAlaThrSerCyAsnAspProGlyIleProGlnAsnGlySerArgSerGlyA      | 1158 |
| Db | 909  | TGTCAACAGCAAGCTCTGCAATACCTGGGATTCGCGAGAAAGGAGTGGAGTGGTG          | 968  |
| Qy | 1158 | apsSerTPGluValaGlyAspSerThrValPheGlnCyAspProGlyTyrAlaLeuGlnG     | 1178 |
| Db | 969  | ACAGTGGGAGAGCGGGGAGCTCCAAGTGTCCAGTGTGACCTTGGGCTGAGCGTGCAGG       | 1028 |
| Qy | 1178 | lysSerAlaGluIleSerCyValIyylleGluAsnArgPheThrPglInProSerProp      | 1198 |
| Db | 1029 | GAAATGCAAGATCAGCTGTGTGAAGATCGAGAACAGGTTCTTCTGCAAGCCGCGCGC        | 1088 |
| Qy | 1198 | roThrCyAlaAlaProCyAspGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerP    | 1218 |
| Db | 1089 | CAACATGATGTGCTCTGCGGGGAGACCTGACAGACACATCTGGAGTCACTCTGCAC         | 1148 |
| Qy | 1218 | roAsnTyrProGluProTyrProProGlyIlysgIuCyAspTPylsValThrValSerP      | 1238 |
| Db | 1149 | CAAAATTACCGAGAACCTTACCAGCGAGAGAGTGTGAGTGAAGAGTACCGTCTCAC         | 1208 |
| Qy | 1238 | roAspTyrValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheL     | 1258 |
| Db | 1209 | CAGACTACGTATCGCCCTGCTATTTAACATCTTAACTGTGAGACCTGTGATATGCTCC       | 1268 |
| Qy | 1258 | eulHisIleTyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerG    | 1278 |
| Db | 1269 | TCCATATCTACAGAGAGAGGAGCTCTTCAGCCCTCTCATAGAGAGCTTCTATGGCTCC       | 1328 |
| Qy | 1278 | InLeuProGlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspA     | 1298 |
| Db | 1329 | AGCTCCAGAGCGGCGATTGAAGAGAGCAACAGCTTCTCTGCTTCCGAGCGCATG           | 1388 |
| Qy | 1298 | lasSerValSerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCyap    | 1318 |
| Db | 1389 | CATCTGTAGCAATGTGGCTTCCGTCAATTGACTTACAGAAACCCCGGAGAGCATGTT        | 1448 |
| Qy | 1318 | heAspProGlySerIleLysAsnGlyThrArgValaGlySerAspLeuLysLeuGlySerS    | 1338 |
| Db | 1449 | TTGATCTCGGTTCCATCAAGAGCGGCAACAGGAGGAGTCCGACTTAAGCTGGGCTCTC       | 1508 |
| Qy | 1338 | erValIthrTyrTyrCyHisIleGlyGlyTyrGluValaGluGlyThrSerThrLeuSerCyai | 1358 |
| Db | 1509 | CCGTCACCTACTACTGCGACAGGGGGCTACAGAAAGTTGAGGGGCACTCGACCTGAGCTCA    | 1568 |
| Qy | 1358 | leLeuGlyProAspGlyLysProValIleTPAsnAsnProArgProValCyethrAlaProC   | 1378 |
| Db | 1569 | TCTGTGGGGCCCTGATGGAGAGCCGCTGTGGAACAATCCCGGCGAGTGTGCAAGCCCTC      | 1628 |
| Qy | 1378 | ysGlyGlyGlnTyrValaGlySerAspGlyValIleuSerProAsnTyrProGlnAsnT      | 1398 |
| Db | 1629 | GTGGGGAGACATATGTGCGTTTCGAGCGAGATGGTCTTGTCTCCCAATACCCAGAACT       | 1688 |
| Qy | 1398 | yrThrSerGlyGlnIleCyLeuTyrPheValThrValProLysAspTyrValIlePheG      | 1418 |
| Db | 1689 | ACACAGTGAAGATGTGCTGTATTTTGTACTGTGCGCCAGAGACTATGTGTGTTTG          | 1748 |
| Qy | 1418 | lyGlnPheAlaPhePheHisThrAlaLeuAsnAspValaGlyValaHisAspGlyHis       | 1438 |
| Db | 1749 | GCCAGTTCGCTTCTTTCACAGCGCCCTCAACAGAGTGTGAGTTCACAGAGCGCCACA        | 1808 |
| Qy | 1438 | erGlnHisSerArgLeuLeuSerSerLeuSerGlySerHisThr                     | 1452 |
| Db | 1809 | GCCAGCACTGCGGCTCTTACGCTCCCTCTCGGGCTCCCATACAGATATCCGGGCTCGG       | 1868 |
| Qy | 1452 | -----  | 1452 |
| Db | 1869 | CCAGTGTGGGATGTGTGGCCGGGGGATCATCGTCCGGCTAAAGAGAGAGCTCTA           | 1928 |
| Qy | 1452 | -----  | 1452 |
| Db | 1929 | GAAAGACCCCATGCGCGCAGAGTGAACCTTACGGCTTGTGTCGTCTGTCTGTGTCG         | 1988 |
| Qy | 1452 | -----  | 1452 |
| Db | 1989 | CTTGTCTACAAAGCAGCAGCAGCTGTAGACTCCAACTACGGGGGCTTACAGAGCT          | 2048 |
| Qy | 1452 | -----  | 1452 |
| Db | 2049 | GCCCTCACCCAGACTGTGTCTACACCGCCCTTGTGTGTAGCTTCTCTGTGTGAATG         | 2108 |
| Qy | 1452 | -----  | 1452 |
| Db | 2109 | GCAACTACACTAATGTGCTGACAGTCCAGTGTGTCTCTCCCTGCGCCATCTGTA           | 2168 |
| Qy | 1452 | -----  | 1452 |
| Db | 2169 | CTGCACCAAGACAAAGATATACCTTGTCTTCTGCTCAAAAGACTGTCACTTACCTTG        | 2228 |
| Qy | 1453 | -----GlyGluSerLeuProLeuAlaThrSerAsnGlnValIleuIleL                | 1467 |
| Db | 2229 | TTTCTGTGCTGCATGACAGAGAACTACTGCTTGTGCACTCAATCAAGTTCTCATTA         | 2288 |
| Qy | 1467 | ysPheSerAlaLysGlyLeuAlaProAlaArgGlyPheHisPheValTyrGln-----       | 1484 |
| Db | 2289 | AGTTCAAGCGCAAGAGGCTTCGCAACAGCGAGAGCTTCCACTTGTCTACCAAGTATGC       | 2348 |
| Qy | 1485 | -----AlaValProArgThrSerAlaThrGlnCySerSerValProG                  | 1499 |
| Db | 2349 | AGGACATGACCGCGGAGCGGTTCTCGAAGCCGAGCGAGTGCAGTGTGTCGGG             | 2408 |
| Qy | 1499 | InProArgTyrGlyLysArgLeuGlySerAspPheSerValaGlyAlaIleValaArgPheG   | 1519 |
| Db | 2409 | AACCCGCTATGCAAGAGCTGGGAGTATCTTCGAGGGGCGCATGTCCGGCTCG             | 2468 |
| Qy | 1519 | IucYAsnSerGlyTyrAlaLeuGlnGlySerProGluIleGluCyLeuProValProG       | 1539 |
| Db | 2469 | AATGCAACTCCGGCTATGCTCTGACAGGGGCGCAGAGATCGAGTGTCTCTCGTGTCTG       | 2528 |
| Qy | 1539 | IylAlaLeuAlaGlnTPAsnValSerAlaProThrCyValaValaProCyGlyGlyAsnL     | 1559 |
| Db | 2529 | GGGCTTGGCCCATATGATGTCTGAGCGCCAGCGTGTGTGTGCGGTGTGAGGAGCAAC        | 2588 |
| Qy | 1559 | eulThrGluArgArgGlyThrIleLeuSerProGlyPheProGluProTyrLeuAsnSerL    | 1579 |
| Db | 2589 | TCAAGAGGCGCAGGGGACACATCTCTCCCTGCTTCCAGAGCGGTACTCAACAGCGC         | 2648 |
| Qy | 1579 | eulAsnCyValITryllyIleValaValaProGluGlyValaGlyIleGlnIleGlnValaLis | 1599 |
| Db | 2649 | TCAACTGTGTGTGAAGATCGTGTGTCCCGAAGCGCTGCAATCCAGATCCAAAGTGTCA       | 2708 |
| Qy | 1599 | erPheValThrGluGlnAsnTPAspSerLeuGluValaPheAspGlyAlaAspAsnThrV     | 1619 |

|    |      |   |      |    |      |  |      |
|----|------|---|------|----|------|--|------|
| Db | 2709 | GTATTGGACAGACAGAACTGGGACTCGCTGGAAGATTGAGTGACAGATAACTG           | 2768 | Db | 3789 | GCCCGTACTCCAGACTCCAGAGACTGTCGTGCTGATCACTCGTGCCCAATTGGCCATGG        | 3848 |
| Qy | 1619 | alThrMetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerA    | 1639 | Qy | 1936 | lyValArgLeuLeuLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleT       | 1956 |
| Db | 2769 | TAAACAAGCTGGGAGATTCTCAAGAAACACCGTGGCTCTTCAACAGCACTCCCA          | 2828 | Db | 3849 | GGCTCGGCTCAACTCACTGCTGTCAGACAGAGCCCTCTGGAATTTCATCAACCATCT          | 3908 |
| Qy | 1639 | snGlnLeuTYrLeuHisPheTYrSerAspIleSerValSerAlaAlaGlyPheHisLeuG    | 1659 | Qy | 1956 | rPAspGlyProGlnGlnThrLapArgLeuGlyValPheThrArgSerMetAlaLeuT          | 1976 |
| Db | 2829 | ACCAAGCTCTACCTCATTTCTACTCAGATACACGATATCTGCAGCTGGCTCCACTGG       | 2888 | Db | 3909 | GGGATGGGCACAGCAACAGCAACACCGCTCGCGCTCTTCAACCGGAGCATGGCCAA           | 3968 |
| Qy | 1659 | luTYrlyeThrValGlyLeuSerSerCySProGluProAlaValProSerAsnGlyValL    | 1679 | Qy | 1976 | yeThrValGlnSerSerSerSerAsnGlnValLeuLeuLybPheHisArgAspAlaIaThrG     | 1996 |
| Db | 2889 | AGTCAAAACGCTGGGCTGAGCAGTTCGCGAACCCTGCTGTCGCCAGTACACGGGCTGA      | 2948 | Db | 3969 | AAACAGTGCAGAGTTTCATCAACCAAGCTCTGCTCAAGTTCCACCGCTGATGACGCACAG       | 4028 |
| Qy | 1679 | yeThrGlyValArgTYrLeuValAsnAspValValSerPheGlnCyGluProGlyTYrA     | 1699 | Qy | 1996 | lyGlyIlePheAlaIleAlaPheSerAlaTYrProLeuThrlycCySProProThrI          | 2016 |
| Db | 2949 | AGACTGGCGAGCGCTACTTGTGTGATATGATGATGTCCTTTCAGTGTGACCCGGGATATG    | 3008 | Db | 4029 | GGGAGATCTTCCCAAGCTTCTCGCTTATCCACTCAACCAATGCCCTCTCCACCA             | 4088 |
| Qy | 1699 | laLeuGlnGlyHisAlaHisIleSerCySmetProGlyTYrThrValArgArgTYrAsnTYrP | 1719 | Qy | 2016 | leLeuProAsnAlaGluValValThrGluAsnGluGluPheAsnIleGlyAspIleValA       | 2036 |
| Db | 3009 | CCCTCCAGGGGCCAGCCCAATCTCTGCATGGCCGGAACAGTGGCCGATGGAACTGACC      | 3068 | Db | 4089 | TCCTCCCAAGCCGGAAGTCGTCAAGAGAAATGAAGAAATTCATTAAGTGACATCGTAC         | 4148 |
| Qy | 1719 | roP-roP-roLeuCySIIeAlaGlnCySgLyGlyThrValGluGluMetGluGlyValIleL  | 1739 | Qy | 2036 | rgTYrArgCySLeuProGlyPheThrLeuValGlyAsnGluIleLeuThrCySlyLeuG        | 2056 |
| Db | 3069 | CTCTCCACTCTGTATTGACACAGTGTGGGGAAACAGTGAAGATGAGAGGGGTGATCC       | 3128 | Db | 4149 | GCTACAGATGCTCTCCCTGGCTTTACTTATGTTAGGGAAATGAATTCGTGACCTGCAACTTG     | 4208 |
| Qy | 1739 | euserProGlyPheProGlyAsnTYrProSerAsnMetAspCySserTYrIleAlaAla     | 1759 | Qy | 2056 | lyTYrThrLeuGlnPheGluGlyProProProlaCyGluValHisCySProThrArg          | 2076 |
| Db | 3129 | TGAGCCCCGGCTTCCAGGCAATACCCCAAGTACATGACCTGCTCTGAAATAATGACAC      | 3188 | Db | 4209 | GAACCTACTCGTAGTTGAGAACCAACCCGATATGGAAGTGACGTCCAAACAAATG            | 4268 |
| Qy | 1759 | euproValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrGluProAsnHisA    | 1779 | Qy | 2076 | luLeuLeuThrAspSerThrGlyValIleLeuSerGlnSerTYrProGlySerTYrProG       | 2096 |
| Db | 3189 | TGCCCGTGGGCTTGGAGCTCACATCCAGTTCTCGAATCTTCCACCCGAGCCCAACACAG     | 3248 | Db | 4269 | AGCTTCTGACAGACTCCACAGGCGTATCTGAGCCAGACAGCTACCTCGAAGCTATCCC         | 4328 |
| Qy | 1779 | spTYrIleGluIleArgAsnGlyProTYrGluThrSerArgMetGlyArgPheSerG       | 1799 | Qy | 2096 | lnPheGlnThrCySserTYrPheThrLeuValArgValGluProAspTYrAsnIleSerLeuTYrV | 2116 |
| Db | 3249 | ACTACATAGAATTCGGAAATGGCCCTATAGACACACCCGATGAGGAAGATTCAAGTG       | 3308 | Db | 4329 | AGTTCCAGACCTGCTTGGCTGGTGGAGAGTGGAGCCGAGCTATVACAATCTCCCTCAAG        | 4388 |
| Qy | 1799 | lySerGlnLeuProSerSerLeuLeuSerThrSerHisGluThrThrValTYrPheHisA    | 1819 | Qy | 2116 | alGluTYrPheLeuSerGlnLySglnTYrAspGluPheGluIlePheAspGlyProSerG       | 2136 |
| Db | 3309 | GAAGCGAGCTTCCAAAGTCTCTCTCTCCAGTCCCAAGACACACCGGTATTTTCCACA       | 3368 | Db | 4389 | TGGAGTACTTCTTCCAGCAGAAAGCAATATGATGATTTGAATTTTGAATGCTCATCG          | 4448 |
| Qy | 1819 | erAspHisSerGlnAsnArgProGlyPheLySLeuGluTYrGln-----               | 1833 | Qy | 2136 | lyGlnSerProLeuLeuLySAlaLeuSerGlyAsnTYrSerAlaProLeuIleValThrs       | 2156 |
| Db | 3369 | GCGACCACTCCCAAGATCGGCCAGAGATTCAAGCTGGAGTATCAGATTGACTTACTCC      | 3428 | Db | 4449 | GACAGAGTCTCTGCTGTAAGGCCCTCAGTGGGAATTATCTAGCTCCCTGATTGACACA         | 4508 |
| Qy | 1833 | -----   | 1833 | Qy | 2156 | erSerSerAsnSerValTYrLeuArgTYrPserSerAspHisAlaTYrAsnArgLySgLyP      | 2176 |
| Db | 3429 | ACCAAGATTCTCTCTTCTGAGAGTTTGATCTCTCGAGTTGAAAGAACCAACTCA          | 3488 | Db | 4509 | GCTCAAGCAACTGTGTACTGCTGGTGGTATCATGATCAAGCTTACCAATCGGAAGGCT         | 4568 |
| Qy | 1834 | -----   | 1837 | Qy | 2176 | heLyIleArgTYrSerAlaProTYrCySserLeuProArgAlaProLeuHisGlyPheI        | 2196 |
| Db | 3489 | CTCTCCCGCTGCGCGCTTCTATGTCTGAGATCTTGATCTTGATCTGTGTAAGCTTATGAA    | 3548 | Db | 4569 | TCAAGATCCGCTATTC-----  | 4585 |
| Qy | 1837 | euglnGluCySProAspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTYrA    | 1857 | Qy | 2196 | leLeuGlyGlnThrSerThrGlnProGlyLySeriIleHisPheGlyCYsAsnAlaGlyT       | 2216 |
| Db | 3549 | TTCAGAAGTGCACAGACCCAGAGCCCTTTCCAATGGCATTTGAGAGGGAGCTGCTACA      | 3608 | Db | 4586 | -----GGCCAGACCGACGCCAGCCGGGGCTCATCTCACTTGGCTGCAACCGCGCT            | 4640 |
| Qy | 1857 | snValGlyGlnSerValThrPheGluCyLeuLeuProGlyTYrGlnLeuThrGlyHisArg   | 1877 | Qy | 2216 | TYrArgLeuValGlyHisSerMetAlaIleCySThrArgHisProGlnGlyTYrHisLeuT      | 2236 |
| Db | 3609 | ACGTGGGACATAGTACCTTGAGTGCCTCCCGGGGTATCATATTGACGTGGCCACCCG       | 3668 | Db | 4641 | ACCGCTGTGGGACACAGATGCGCATCTTACCCGGCACCCCAAGGCTTACCACTGT            | 4700 |
| Qy | 1877 | alLeuThrCySglnHisGlyTYrAsnArgAsnTYrAspHisProLeuProLyCySgLuV     | 1897 | Qy | 2236 | rPSerGluAlaIleProLeuCySglnAlaLeuSerCySgLyLeuProGluIleLapPolySA     | 2256 |
| Db | 3669 | TCCTCAAGTGTCAACATGGACCAACCGAACTGGGACCAACCCCTGCCCAAGTGTGAAG      | 3728 | Db | 4701 | GGAGGAGACCATCTCTCTGTCAAGCTCTTCTCTGTGGCTTCTGAGGCCCCCAAGA            | 4760 |
| Qy | 1897 | alProCySgLyGlyAsnIleThrSerSerAsnGlyTYrValTYrSerProGlyPheProS    | 1917 | Qy | 2256 | snGlyMetValPheGlyLySgLuTYrThrValGlyTYrlyAsaIleValTYrSerCySerg      | 2276 |
| Db | 3729 | TCCCTTGTGGCGGGAACATCATCTTCCAAACGGCACTGTGTACTCCCGGGGTTCCCTA      | 3788 | Db | 4761 | ATGGAATGTGTGGCAAGAGATACACAGTGGGAAACCAAGGCCATGTACGTGCACTG           | 4820 |
| Qy | 1917 | erProTYrSerSerSerGlnAspCySValTYrLeuIleThrVal--ProIleGlyHisG     | 1936 | Qy | 2276 | luGlyTYrHisLeuGlnAlaGlyValaGluAlaThrAlaGluCySLeuAspThrGlyLeuT      | 2296 |
| Db | 1917 | erProTYrSerSerSerGlnAspCySValTYrLeuIleThrVal--ProIleGlyHisG     | 1936 | Db | 4821 | AAGGCTACACCTCAGGACAGCGCTGAGGCCACTGCAAGATGTCTGACACAGGCTTAT          | 4880 |

QY 2296 rPserAsnArgAsnValProProGlnCysVal-----P 2307  
 Db 4681 GGAGCAACCGCAATGTCCCAACAGATGTCTCGGAGTCTCTCGGGCAATGAGCGGGT 4940  
 QY 2307 roValThrCysProAspValSerSerIleSerValGluHisGlyArgTrpArgLeuIleP 2327  
 Db 4941 CTGTGACTGTTCCTGATGTCTAGTACATCAGCGTGAAGACATGCGCATGAGCGCTTATCT 5000  
 QY 2327 heGluThrGlnTrpGlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTrpTrpT 2347  
 Db 5001 TTGAGACACAGTATCAGTTCCAGGCGCCAGTGAATGCTCATGTGACCTTGCTCACTACT 5060  
 QY 2347 YrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLeuTrpSerLeuGlyAspSerT 2367  
 Db 5061 ATACTGGCCAAAGGCTCATCCGCTGTCCAGGCCAATGGAATGAGCTCGGGAGCTCTA 5120  
 QY 2367 hrProThrCysArgIleIleSerCysGlyGluLeuProIleProProAsnGlyHisArgI 2387  
 Db 5121 CGCCCACTGCGCATCATCTCTGTGAGAGCTCCGATTCGCCCAATGGCCACCGCA 5180  
 QY 2387 leglyThrLeuSerValTrpGlyAlaThrAlaIlePheSerCysAsnSerGlyTrpTrL 2407  
 Db 5181 TCGAACAACGTGTCTGTCTACGGGGCACAGCCATCTTCTCCGCAATCCGATACACAC 5240  
 QY 2407 euValGlySerArgValArgIuCyMetAlaAsnGlyLeuTrpSerGlySerGlyValA 2427  
 Db 5241 TGGTGGCTCCAGGGGTGGTAGTGATGCGCAATGGGCTGTGAGTGGCTCTGATGATCC 5300  
 QY 2427 rGcYsLeu----- 2429  
 Db 5301 GCTGCTTGGCACTCAGACCAAGCTCCATCTTCTATAGCTCTCTGATGTAC 5360  
 QY 2430 -----AlaGlyHisCysGlyThrProGluProIleValAsnG 2442  
 Db 5361 TCTCTTCCCATTCCTCAACCAAGCTGAGACCTGTGGAGCTCTGAGGCCATTTGTCAACG 5420  
 QY 2442 LysHisIleAsnGlyGluAsnTrpSerTrpArgGlySerValValTrpGlnCysAsnAlaG 2462  
 Db 5421 GACACATCAATGGGAGAACTACAGCTACCGGGGAGGTGTGTGATCCAAATGCAATGTGTG 5480  
 QY 2462 LysPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisHisTrpSerGlyL 2482  
 Db 5481 GCTTCGGCTGATGGCGATGTCTGTGGCATCTGCGACAGATCATACGTGTGGGCA 5540  
 QY 2482 YsThrProPheCysVal----- 2487  
 Db 5541 AGACCCCTTCTGTGTGATGTAAAGCAGATTGCTGCTGCTGCTGTGTGTG 5600  
 QY 2488 -----ProIleThrCysGlyHisAspProGlyAsnP 2497  
 Db 5601 ATG 5660  
 QY 2497 roValAsnGlyLeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValLysPheValC 2517  
 Db 5661 CTGTCAAGGGCTCTCACTAGGGTAAACAGTTTAACCTTAAGATGTGTCAAGTTTGT 5720  
 QY 2517 YsaAsnProGlyTrpMetAlaGlnGlyAlaAlaArgSerGlnCysLeuAlaSerGlyGlnT 2537  
 Db 5721 GCAACCCCTGGGTATATGCTGAGGGGGCTGCTAGGTGCCAATGCTGCGCAGCGGCAAT 5780  
 QY 2537 rPserAspMetLeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGlnUA 2557  
 Db 5781 GGAGTGAATCTGTGCCACCTGCGAATATCATCAATGTACAGATCTTGGACACCAAGAA 5840  
 QY 2557 anSerValArgGlnValHisAlaSerGlyProHisArgPheSerPheGlyThrThrValS 2577  
 Db 5841 ATAGTGTCTGATAGGTCCAAGCCGAGCGCGCCAGAGGTTCAAGCTTGGCACCACCTGTGT 5900  
 QY 2577 erTrpArgCysAsnHisGlyPheTrpLeuLeuGlyThrProValLeuSerCysGlnGlyA 2597  
 Db 5901 CTTACCGGTGCAACCAAGCGCTTCTTACTCTCTGGGACACCCAGTGTCTGCTGCGAGGAG 5960

QY 2597 spGlyThrTrpAspArgProArgProGlnCysLeu 2608  
 Db 5961 ATGGCATGTGGAAACCGTCCCGCCCGCCCAAGTGTCTC 5995  
 RESULT 4  
 US-10-276-934-5  
 ; Sequence 5, Application US/10276934  
 ; Publication No. US20030180750A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: University of Leeds  
 ; APPLICANT: Markham, Alexander F.  
 ; APPLICANT: Jackson, Andrew P.  
 ; APPLICANT: Woods, Christopher G.  
 ; TITLE OF INVENTION: Treatment of Cancer and Neurological Diseases  
 ; FILE REFERENCE: 9052-144  
 ; CURRENT APPLICATION NUMBER: US/10/276,934  
 ; CURRENT FILING DATE: 2002-11-20  
 ; PRIOR APPLICATION NUMBER: PCT/GB01/02240  
 ; PRIOR FILING DATE: 2001-05-21  
 ; PRIOR APPLICATION NUMBER: GB0012186.3  
 ; PRIOR FILING DATE: 2000-05-20  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 5  
 ; LENGTH: 7323  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (34)..(34)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (128)..(128)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (638)..(638)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1502)..(1502)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1538)..(1538)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1586)..(1586)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1782)..(1782)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1787)..(1787)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (2195)..(2195)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (2213)..(2213)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (3179)..(3179)  
 ; OTHER INFORMATION: "n" is any nucleotide  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature



|    |      |   |      |
|----|------|---|------|
| Qy | 640  | AsnAspPheTyrValGlyAlaLeuValThrPheSerCyAspSerGlyTyrThrLeuSer     | 659  |
| Db | 2799 | GGAGACTTTGGCATCAGGCTCCACAGATGACTTTCAGCTCTGACCCGGGTACACACTAAGT   | 2858 |
| Qy | 660  | AspGlyGluProLeuGluCyGluGluProAsnPheGlnTyrPheArgAlaLeuProSerCyAs | 679  |
| Db | 2859 | GACACAGAGCCCTCTGCTGTGTGAGAGAAACCACAGTGAACACAGCTTGGCCAGCTGC      | 2918 |
| Qy | 680  | GluAlaLeuCyGlyGlyPheIleGlnIlySerSerGlyThrIleLeuSerProGlyPhe     | 699  |
| Db | 2919 | GAGCGTATATGGAGGCTACATCCAGAGGAAGGTGAAACAGTCTTTCTCTCGGGATT        | 2978 |
| Qy | 700  | ProAspPheTyrProAsnAsnLeuAsnCySthrTrpIleIleGluThrSerHisGlyAs     | 719  |
| Db | 2979 | CCAGATTTTATTCMAACTCTTAACTGACAGTGGACCATTTGAAGTCTCATGGGAAA        | 3038 |
| Qy | 720  | GlyValPhePheThrPheHisThrPheHisLeuGluSerGlyHisAspTyrLeuLeuIle    | 739  |
| Db | 3039 | GGAGTTCAATGATCTTTTACACCTTTATCTTGAGATTTCCACGACTTATTTACTGATC      | 3098 |
| Qy | 740  | ThrGluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAla    | 759  |
| Db | 3099 | ACAGAGGATGGAAGTTTTTCCGAGCCGTTGGCAGGCTCACCGGGTGGTGTGCTCAT        | 3158 |
| Qy | 760  | ProIleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPhe    | 779  |
| Db | 3159 | ACGATTCAGAGGACGGCTCTTGGGAACCTTCACTGCCACGCTTGGTTTATTCAGACTTC     | 3218 |
| Qy | 780  | SerIleSerTyrGlnGlyPheAsnIleThrPheSerGlyTyrAspLeuGluProCyGluGlu  | 799  |
| Db | 3219 | TCAATTTCGACGAGGGCTCTCAATATTCATTTTCAGAAATAGACTGGAGCATGATAT       | 3278 |
| Qy | 800  | GluProGluValProAlaTyrSerIleArgIlyGlyLeuGlnPheGlyValGlyAspThr    | 819  |
| Db | 3279 | GATCTGGAGTCCCTGCTTACAGCCGGAAGAAATGGTTTTCACTTGGTGGGAGACTCT       | 3338 |
| Qy | 820  | LeuThrPheSerCyAspPheProGlyTyrArgLeuGlnGlyThrAlaArgIleThrCyLeu   | 839  |
| Db | 3339 | CTGACGTTTTCTGCTCTCTGGGAATGCTTAGAAGGTGCACCACTTAACTGCTGCTG        | 3398 |
| Qy | 840  | GlyGlyArgAspArgLeuThrPheSerProLeuProArgCyValAlaGluCyGlyValAsn   | 859  |
| Db | 3399 | GGTGGGGCCCGCGGTGTGGAGTGACCTTGCCAAAGTGTGGCGGAATGTGGACCA          | 3458 |
| Qy | 860  | SerValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTyrAsnAsn    | 879  |
| Db | 3459 | AGTGCAGAGGAATGAGAGAACNTATGCTCTCAATTTTCCATTCCAATTATGATAT         | 3518 |
| Qy | 880  | AsnHisGluCyValIleTyrSerIleGlnThrGlnProGlyIlyGlyIleGlnLeuIlyAla  | 899  |
| Db | 3519 | AACCATGATGTATCTATATAATATGAAACAGAACCGGACAGGGCATCACCTTAGAAC       | 3578 |
| Qy | 900  | ArgAlaPheGluLeuSerGlnGlyAspValLeuIlyValIlyAspGlyAsnAsnAsnSer    | 919  |
| Db | 3579 | CGAAGCTTCCAGCTGTGTGAAGAGATATCTTAAAGGATATGATGAGAAAAGACAGTTC      | 3638 |
| Qy | 920  | AlaIArgLeuLeuGlyValPheSerHisSerGluMetGlyValThrLeuAsnSerThr      | 939  |
| Db | 3639 | TCAGTGCATCGGGACGCTCAGCTAAATAATGAATCTCGGGGCTGATCTTAAACAGACA      | 3698 |
| Qy | 940  | SerSerSerLeuTrpLeuAspPheIleThrAspAlaGluAnthrSerIlyGlyPheGlu     | 959  |
| Db | 3699 | TCCATCATCCTGTGGCTAGAGTTCMAACCAATGATCTGACACCGACCAAGTTTTCAA       | 3758 |
| Qy | 960  | LeuHisPheSerSerPheGluLeuLeuIleIlyCyGlyAspProGlyThrProIlyPheGly  | 979  |
| Db | 3759 | CTCACCTTATCCAGTTTGTATCTGTATAAAATGTAGATCCGGGGATCCCTAATCAAGGC     | 3818 |
| Qy | 980  | TyrIlyValHisAspGlnGlyHisPheAlaGlySerValSerPheSerCyAspPro        | 999  |
| Db | 3819 | TATAGGATCCGATGAGAGGACACTTTACCGACACTGATGTTCTGTACAGTTGCAACCG      | 3878 |
| Qy | 1000 | GlyTyrSerLeuArgIlySerGlnIleuLeuCyLeuSerGlyGluArgThrTrp          | 1019 |

|    |      |  |      |
|----|------|--|------|
| Dd | 3879 | GGGTRACGCATGCAATGGAGCAACACCTTGACCTGTTGATGGAGACGAGAGTGTGG         | 3938 |
| Qy | 1020 | AspArgProLeuProThrCysValAlaGluCysGlyGlyThrValAlaGlyValSer        | 1039 |
| Dd | 3939 | GACAACACCTACTCTCTCGTATGACCGGAATGTGTGTGCATGCATGCACGCACATCA        | 3998 |
| Qy | 1040 | GlyGlnValLeuSerProGlyTYrProAlaProTyrGlnHisAsnLeuAsnCysIleTrp     | 1059 |
| Dd | 3999 | GGACGAATATGTGCTCCGCTGCTATCCAGCTCCGATATGACAACAACCTCCACTGCCTGG     | 4058 |
| Qy | 1060 | ThrIleGlnIleGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGlu     | 1079 |
| Dd | 4059 | ATTATATAGGAGAGACCCAGGAAGACCATTAAGCTTCATTTGTTTTCAGACGGAG          | 4118 |
| Qy | 1080 | GluValHisAspValLeuArgIleTrpAspGlyProValGluSerGlyValLeuLeuLys     | 1099 |
| Dd | 4119 | ATGGCTACGACATCTCTCAAGTCTGGGACGGCGCGGTGACACATGCATCTCTGAGG         | 4178 |
| Qy | 1100 | GluLeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValLeu        | 1119 |
| Dd | 4179 | GAGTGGAGTGGCTCCGCCCTTCGGAGAGCATTCACACGACCTTCAACTCACCCTCG         | 4238 |
| Qy | 1120 | GlnPheSerThrAspPhePheThrSerLysGlnIlyPheAlaIleGlnPheSerValSer     | 1139 |
| Dd | 4239 | CAGTTTCACACGACTTCTTCATCAGCAAGTCTGCTTCTCATCCAGTTCTCACCTCA         | 4298 |
| Qy | 1140 | ThrIleAlaThrSerCysAsnAspProGlyVilEProGlnAsnGlySerArgSerGlyAspSer | 1159 |
| Dd | 4299 | ATTGCACGCACCTGTAAAGCATCCAGATATGCCCCAAATGGCACCCGCTATGGAGACAGC     | 4358 |
| Qy | 1160 | TrpGlnIleGlyAspSerThrValPheGlnCysAspProGlyTYrAlaLeuGlnGlySer     | 1179 |
| Dd | 4359 | AGAGAGGTGGAGACACCGCATCTTCACAGTGTACCTGGCTATTCAGCTCCAGGACAA        | 4418 |
| Qy | 1180 | AlaGlnIleSerCysValLysIleGlnAsnArgPheThrProGlnProSerProProThr     | 1199 |
| Dd | 4419 | GCCAAATACACTGTGTGCACGTAAATTAACCGGTCTTTTGGCAACAGACCTTCCTACA       | 4478 |
| Qy | 1200 | CysIleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsn     | 1219 |
| Dd | 4479 | TGCATAGCTGCTGTGAGGGAATCTGACGGGCCACAGAGTATTTGTCTACCCAC            | 4538 |
| Qy | 1220 | TyrProGlnProTyrProProGlyLysGlnCysAspTrpLysValThrValSerProAsp     | 1239 |
| Dd | 4539 | TACCACACGCGTATCTCTCGTGGAGGAATGTACTCGAGAGATTAATAAGTGAACCCGGAC     | 4598 |
| Qy | 1240 | TyrValIleAlaLeuValPheAsnIlePheAsnLeuGlnProGlyTYrAspPheLeuHis     | 1259 |
| Dd | 4599 | TTTGTCTACGCGCTTGATTAATCAAAAGTTTCMACTGACGCCACGACTATGACTTCTTCAC    | 4658 |
| Qy | 1260 | IleTYrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTYrGlySerGlnLeu     | 1279 |
| Dd | 4659 | ATCTATGAGGGGAAGATTCACACAGCCCTCATTTGGAGTTTCCAGGGCTCTCAGGCC        | 4718 |
| Qy | 1280 | ProGlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSer     | 1299 |
| Dd | 4719 | CCAGAAAGATATAGAGATGACGGAACAGCGCTGTTCTGGCATTTCCGAGTATGATGCTCC     | 4778 |
| Qy | 1300 | ValSerAsnAlaGlyPheValIleAspTYrThrGlnAsnProAlaGluSerCysPheAsp     | 1319 |
| Dd | 4779 | GTTGGCGCTTTCAGGGTTTCGCCATTAATTAAGAGAAACCCACGGAGAGCTGTGTTGAC      | 4838 |
| Qy | 1320 | ProGlySerIleLysAsnGlyThrArgValGlySerAspLeuLysGlySerSerVal        | 1339 |
| Dd | 4839 | CCAGGAATATATAGAAATGGACACAGAGTTGGAAACAAGCTTCAGACTGGCTGCACCATC     | 4898 |
| Qy | 1340 | ThrTYrTYrCysHisIleGlyTYrGlnValGlnGlyThrSerThrLeuSerCysIleLeu     | 1359 |
| Dd | 4899 | ACCTATCCAGTGTGACTCGGTATATTAAGATTCTTGAACCCCTCATCATCACCTGTGATTT    | 4958 |
| Qy | 1360 | GlyProAspGlyLysProValITrpAsnAsnProArgProValCysThrAlaProCysGly    | 1379 |

Db 4959 GGGGCTGATGGAAACCTCTCGGGACCAAGTCGCTCTCTGCAATGCTCCCTGGGA 5018  
 Qy 1380 GlyIuInTyValIglYserAspGlyValIleuSerProAsnTyrrProGlnAsnTyrrThr 1399  
 Db 5019 GGGCAGTACACGGGATCAGAAAGGGGATGTTTATCACCAAACTACCCCAATATTCACCA 5078  
 Qy 1400 SerGlyInIleCyseLeuTyrrPheValThrValProIysAspTyrrValIlePheGlyGln 1419  
 Db 5079 GCTGTCATAATATGCTCTATTCATCAACGGTACCAAGGAATTCGATGCTTTGGACAG 5138  
 Qy 1420 PheAlaPhePheHsrThrAlaLeuAspValValGluValHisAspGlyHisSerGln 1439  
 Db 5139 TTTGGCCATTTTCCAGACAGCCCTGAATGATTTGGCCAAATTAATTGTGAGAACCCATGCA 5198  
 Qy 1440 HisSerArgLeuSerSerLeuSerGlySerHisThrGlyInIleuSerLeuProAla 1459  
 Db 5199 CAGCCAGACTTCTCACTCACTCTCGGGGCTCAGCAGGGAACAAATGGCCCTGGCT 5258  
 Qy 1460 ThrSerAsnGlnValLeuIleValPheSerAlaIleValLeuAlaProAlaArgGlyPhe 1479  
 Db 5259 ACGTCAATCAAAATTCGTCTCCGATTCAGTCAAGAAAGCGGTGCTCTGCGCGGCTTC 5318  
 Qy 1480 HisPheValTyrrGlnAlaValProArgThrSerAlaThrGlnCyseSerValProGln 1499  
 Db 5319 CACTTCGTATCAAGCTGTTCTCGTACCAAGTACACCCAAATGCACTCTGCTCCCGAG 5378  
 Qy 1500 ProArgTyrrGlyValArgLeuGlySerAspPheSerValGlyValAlaIleValArgPheGln 1519  
 Db 5379 CCAGATACGAAAGGAGAAATTTGTTCTGAGTTTCTGCGCGCTCCACATCGTCCATTCAG 5438  
 Qy 1520 CysAsnSerGlyTyrrAlaLeuGlnGlySerProGlnIleGluCyseLeuProValProGly 1539  
 Db 5439 TGCACCCGGGATACCTGCTTCAAGGGTTCACAGCGCGCTCCACATCGTCCATTCAG 5498  
 Qy 1540 AlaLeuAlaGlnInThrPheAsnValSerAlaProThrCyseValValProCyseGlyValAsnLeu 1559  
 Db 5499 GCTTGGCAGAGTGAACGACAGCATCCCGACGTGTGTGTATCCCTGCAAGTGGCAATTTTC 5558  
 Qy 1560 ThrGluArgArgGlyThrIleLeuSerProGlyPheProGlnProTyrrLeuAsnSerLeu 1579  
 Db 5559 ACTCAAGAAAGAGTACATCTCTGCTCCCGGCTACCTGAGCCATACGGAACAACCTTG 5618  
 Qy 1580 AsnCyseValTyrrIleValIleValProGlnGlyAlaGlyIleGlnIleGlnValIleValSer 1599  
 Db 5619 AACTGATATGAGAAATCATAGTTACGAGAGGCTCGGGAATTCAGATCCAAAGTACAT 5678  
 Qy 1600 PheValThrGluGlnAsnThrPheSerLeuGlnValPheAspGlyAlaAspAsnThrVal 1619  
 Db 5679 TTTGCCACGAGAGCAAGTGGAGCTCCCTTGAGATCCAGATGTGGGATGTGACCGCA 5738  
 Qy 1620 ThrMetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerAsn 1639  
 Db 5739 CCAGACTGGGAAGCTTCTCAGGCAACAGTACCGGCACTGCTGAACAGTACTTCCAAC 5798  
 Qy 1640 GlnLeuTyrrLeuHisPheTyrrSerAspIleSerValSerAlaAlaGlyPheHisIleuGln 1659  
 Db 5799 CAACTTACCTGATTTCCAGTCTGACATTAAGTGGCAGCTGCTGTTTCCACCTGGAA 5858  
 Qy 1660 TyrrLeuThrValGlyLeuSerSerCyrrProGlnProAlaValProSerAsnGlyValIleVal 1679  
 Db 5859 TAAATAACGTAGTGTCTGTGATGCAACCAAGAACCGCCCTCCCAACGACATCAAA 5918  
 Qy 1680 ThrGlyIleArgTyrrLeuValAsnAspValIleValSerPheGlnCyseGlyProGlyTyrrAla 1699  
 Db 5919 ATCGAGATTCGATKATGTGGAAACAGCTGCTCTCTTCAAGTGGACCGCGGATACCC 5978  
 Qy 1700 LeuGlnGlyHisAlaHisIleSerCyrrMetProGlyThrValArgArgTyrrAsnTyrrPro 1719  
 Db 5979 CTGCAAGGCGGTTCACCACTTCTGTATGCCAGGAGCGGCTTCGCGGTTGGAACATATCG 6038  
 Qy 1720 ProProLeuCyrrIleAlaGlnCyseGlyValThrValGluGluMetGluGlyValIleLeu 1739  
 Db 6039 TCTCCCTGTGATTCACCTGTGAGGAGCGCTGAGCACTTGGTGTGTGTATCTCG 6098

Qy 1740 SerProGlyPheProGlyAsnTyrrProSerAsnMetAspCyseSerTyrrIleAlaLeu 1759  
 Db 6099 AGCCCCGGCTTCCAGGTTCTTACCCCAACCACTTACGCTGACCTGGAGATCTCATTA 6158  
 Qy 1760 ProValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrGlyProAsnHisAsp 1779  
 Db 6159 CCATTCGGGTATAGTGCACATTTTCAAGTTTTCGATTTTTCACGGAAGCTAATCATGAC 6218  
 Qy 1780 TyrrIleGluIleArgAsnGlyProTyrrGluThrSerArgMetGlyValArgPheSerGly 1799  
 Db 6219 TTCCTTGAATCAAAATGAGACCTTACCAACAGCCCATGATGTGACAAATTTAGCGGC 6278  
 Qy 1800 SerGluLeuProSerSerLeuLeuSerThrSerHisGlyThrThrValTyrrPheHisSer 1819  
 Db 6279 ACGGATCTCCCGCGGCTCTGTGAGCAACAGCATGAAACCTTCATCCACTTTATAGT 6338  
 Qy 1820 AspHisSerGlnAsnArgProGlyPheIleuGlnGlyTyrrGlnAlaTyrrGlyIleuGlnGlu 1839  
 Db 6339 GACCAATTCGCAAAACCGGCAAGATTTAACTTGTCTTACCAAGCTTATGAATTACAGAAC 6398  
 Qy 1840 CysProAspProGlnProPheAlaAsnGlyIleValArgGlyAlaGlyTyrrAsnValGly 1859  
 Db 6399 TGTCCAGATCCACCCCATTTCAAGATGGGTACATGATCAACTCGGATTTACAGCTGGGG 6458  
 Qy 1860 GlnSerValThrPheGluCyseLeuProGlyTyrrGlnLeuThrGlnHisProValLeuThr 1879  
 Db 6459 CAATCAATGATCTTTCAGAGTATATCTGGGTACATTTCAATGAGCCATCTGTCTCACT 6518  
 Qy 1880 CysGlnHisGlyTyrrAsnArgAsnTyrrAspHisProLeuProIleCysGlyValProCys 1899  
 Db 6519 TGTGAGCATGGGATCAAGAACTGAACTGAACTTTTCCAAAGTGTGATGCCCTTGT 6578  
 Qy 1900 GlyIleAsnIleThrSerSerAsnGlyTyrrValTyrrSerProGlyPheProSerProTyrr 1919  
 Db 6579 GGGTCAACAGTAACTTCTCAGAACCGGCAACATCTACTCCCTGGCTTCTCATGATGAT 6638  
 Qy 1920 SerSerSerGlnAspCyseValTyrrLeuIleThrValProIleGlyHisGlyValArgLeu 1939  
 Db 6639 CCGATCTCGAAGAGCTGATTTGGCTATCACAGGTGCTCCAGGGCAGAGATTATCATC 6698  
 Qy 1940 AsnLeuSerLeuLeuGlnInThrGluProSerGlyAspPheIleThrIleTyrrAspGlyPro 1959  
 Db 6699 AACTTCAACCTGTTACAGACGGAAGCTGTCAACGATTCATGCTGTTTGGGAGCGTCTC 6758  
 Qy 1960 GlnGlnInThrAlaProArgLeuGlyValPheThrArgSerMetAlaValIleValGln 1979  
 Db 6759 GATCAGAACTCAACCCAGCTGGAGTTTCAGTGGCAACAGGCCCTCGAAACGCGCTAT 6818  
 Qy 1980 SerSerSerAsnGlnValLeuLeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1999  
 Db 6819 AGCTTCAACCAACAGATCTCTCTCAAGTTCCACAGGCACTTTTCAATAGAGGCTTCTT 6878  
 Qy 2000 AlaIleAlaPheSerAlaTyrr-----ProLeuThrIle----- 2010  
 Db 6879 GTCTCAATTTCCAGCGCACTGATTTTCACTCGTTAGTTAAGACTGAGAATTCATG 6938  
 Qy 2011 TGGTGTATTCAGAGTGTGTCACAGCCTTGTTCAGAGTGAAGTTTCTTATTCAGCC 6998  
 Db 2021 GluValValThrGluAsn 2026  
 Db 6999 GAGGCGGTGATGATTTCT 7016

# RESULT 5

US-10-276-934-6  
 ; Sequence 6, Application US/10276934

; Publication No. US20030180750A1

; GENERAL INFORMATION:

; APPLICANT: University of Leeds

; APPLICANT: Jackson, Alexander P.

; APPLICANT: Jackson, Andrew P.